

2

THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of: Naoki MATSUOKA et al.

Filed : Concurrently herewith

For : SCHEDULING CONTROL SYSTEM AND SWITCH

Serial No. : Concurrently herewith

December 14, 1999

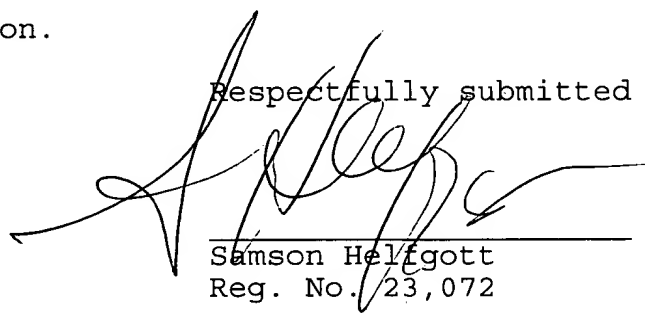
Assistant Commissioner of Patents
Washington, D.C. 20231

SUBMISSION OF PRIORITY DOCUMENT

S I R:

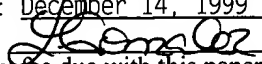
Attached herewith is Japanese patent application No.
10-355888 of December 15, 1998 whose priority has been claimed
in the present application.

Respectfully submitted



Samson Helfgott
Reg. No. 23,072

HELFGOTT & KARAS, P.C.
60th FLOOR
EMPIRE STATE BUILDING
NEW YORK, NY 10118
DOCKET NO.: FUJY16.847
LHH:priority

Filed Via Express Mail
Rec. No.: EL522352918US
On: December 14, 1999
By 
Any fee due with this paper, not fully
Covered by an enclosed check, may be
Charged on Deposit Acct. No. 08-1634

日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application:

1998年12月15日

出願番号
Application Number:

平成10年特許願第355888号

出願人
Applicant(s):

富士通株式会社



CERTIFIED COPY OF
PRIORITY DOCUMENT

1999年 7月15日

特許庁長官
Commissioner,
Patent Office

伴佐山建



出証番号 出証特平11-3050548

【書類名】 特許願

【整理番号】 9805550

【提出日】 平成10年12月15日

【あて先】 特許庁長官殿

【国際特許分類】 H04L 12/50
H04Q 3/495

【発明の名称】 スケジューリング制御装置および交換機

【請求項の数】 23

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

 【氏名】 松岡 直樹

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

 【氏名】 瓦井 健一

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

 【氏名】 朝永 博

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

 【氏名】 加藤 次雄

【特許出願人】

 【識別番号】 000005223

 【氏名又は名称】 富士通株式会社

【代理人】

 【識別番号】 100089244

【弁理士】

【氏名又は名称】 遠山 勉

【選任した代理人】

【識別番号】 100090516

【弁理士】

【氏名又は名称】 松倉 秀実

【連絡先】 03-3669-6571

【手数料の表示】

【予納台帳番号】 012092

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9705606

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 スケジューリング制御装置および交換機

【特許請求の範囲】

【請求項 1】 所望の出力回線に対するスケジューリング対象の各入力回線の送出要求情報を保持する要求情報管理部と、

スケジューリング開始入力回線を示すハイウェイ間ポインタを保持するハイウェイ間ポインタ制御部と、

各入力回線に対応した前記送出要求情報において、検索開始出力回線を示すハイウェイ内ポインタを保持するハイウェイ内ポインタ制御部と、

所望の出力回線に対する送出要求情報を保持する要求管理制御部と、

複数の前記送出要求情報の中からハイウェイ内ポインタ制御部が示す出力回線から検索を開始し、他の入力回線に選択されていない出力回線を選択するスケジューリング処理部とからなり、

前記スケジューリング処理部は、スケジューリング終了後のハイウェイ間ポインタを隣接回線に更新するときに、スケジューリング周期毎に、順方向の隣接回線へ更新する処理と、逆方向の隣接回線へ更新する処理を交互に行うスケジューリング制御装置。

【請求項 2】 所望の出力回線に対するスケジューリング対象の各入力回線の送出要求情報を保持する要求情報管理部と、

スケジューリング開始入力回線を示すハイウェイ間ポインタを保持するハイウェイ間ポインタ制御部と、

各入力回線に対応した前記送出要求情報において、検索開始出力回線を示すハイウェイ内ポインタを保持するハイウェイ内ポインタ制御部と、

所望の出力回線に対する送出要求情報を保持する要求管理制御部と、

複数の前記送出要求情報の中からハイウェイ内ポインタ制御部が示す出力回線から検索を開始し、他の入力回線に選択されていない出力回線を選択するスケジューリング処理部とからなり、

前記スケジューリング処理部は、さらに、

スケジューリング開始回線番号の異なるハイウェイ間ポインタを有する複数の

パイプライン処理手段を有し、

前記ハイウェイ内ポインタ制御部は、前記パイプライン処理手段毎に独立に制御されるスケジューリング制御装置。

【請求項 3】 所望の出力回線に対するスケジューリング対象の各入力回線の送出要求情報を保持する要求情報管理部と、

スケジューリング開始入力回線を示すハイウェイ間ポインタを保持するハイウェイ間ポインタ制御部と、

各入力回線に対応した前記送出要求情報において、検索開始出力回線を示すハイウェイ内ポインタを保持するハイウェイ内ポインタ制御部と、

所望の出力回線に対する送出要求情報を保持する要求管理制御部と、

複数の前記送出要求情報の中からハイウェイ内ポインタ制御部が示す出力回線から検索を開始し、他の入力回線に選択されていない出力回線を選択するスケジューリング処理部とからなり、

前記スケジューリング処理部は、さらに、一定周期内の到着パケット数を各入力回線の論理送出要求情報毎に計測する負荷観測部を有しており、当該負荷観測部の計測数に応じて次の周期のスケジューリングを行うスケジューリング制御装置。

【請求項 4】 前記請求項 3 において、前記スケジューリング処理部は、一定周期内の各出力回線への送出許可パケット数を、前周期の各論理送出要求情報への到着パケット数に制限するスケジューリング制御装置。

【請求項 5】 前記請求項 3 において、前記スケジューリング処理部は、一定周期内のスケジューリングにおける最優先出力回線となる割合を前周期の到着パケット数に応じて決定するスケジューリング制御装置。

【請求項 6】 前記請求項 5 において、前記スケジューリング処理部は、前記ハイウェイ内ポインタとは別に、前周期の到着パケット数に応じて最優先で送出をする回線を決定する最優先送出ポインタを保持し、前周期の入力回線に到着した全到着パケット数分は前記最優先送出ポインタにしたがって各送出要求情報への到着パケット数の割合に応じて出力回線を決定し、前記全到着パケット数分以外については通常のハイウェイ内ポインタにしたがってスケジューリングを実

行するスケジューリング制御装置。

【請求項 7】 前記請求項 3, 4 または 5 において、前記スケジューリング処理部は、現周期内で前周期の到着パケット数分の送出が行われなかったときに、残りの到着パケット数を次の周期の到着パケット数に繰り越す処理を行うスケジューリング制御装置。

【請求項 8】 前記請求項 3 において、前記スケジューリング処理部は、一定周期内のスケジューリングにおける最優先出力回線となる割合を、前周期の送出要求情報長に応じて決定するスケジューリング制御装置。

【請求項 9】 前記請求項 3, 4 または 5 において、各入力回線毎のスケジューリング処理部と、2 個の入出力ポートと、外部入出力と内部のリング接続を切り替えるセレクタ部と、拡張出力への送出遅延量が変更可能な遅延部とを備えた拡張ユニットを形成し、この拡張ユニットをリング接続したスケジューリング制御装置。

【請求項 10】 送出要求情報とともに、サービス品質クラスをスケジューリング処理部に通知する手段と、

第 1 のスケジューリング時に各入力回線の第 1 優先クラスの送出要求情報をもとにスケジューリング処理を行うスケジューリング処理部と、

第 2 のスケジューリング時において、前記第 1 のスケジューリングで出力回線が確定しておらず、かつ第 2 優先クラスの要求情報をもとにスケジューリングを実行するスケジューリング処理部とからなり、

前記スケジューリング処理部は、2 以上のクラスの中の高優先クラスからスケジューリングを行うスケジューラと、このスケジューラで決定した出力回線およびグループ内の帯域制御を行うバッファとを備えたスケジューリング制御装置。

【請求項 11】 前記請求項 10 において、前記スケジューリング処理部はさらに、

前記第 1 優先クラスのスケジューリング後に、バッファへ結果を通知する通知手段と、

前記スケジューラで決定された回線内のサービス品質クラスの帯域および遅延

優先制御を行う前記バッファの帯域制御部とを備え、

前記サービス品質クラスの所定の帯域内であれば、前記バッファはスケジューラからの結果を受け入れて当該バッファに蓄積されたデータを送出し、帯域外であれば、その送出要求情報を無効としてスケジューラに通知する、このようにして第2のスケジューリングを行うスケジューリング制御装置。

【請求項 12】 送出要求情報とともに、サービス品質クラスをスケジューリング処理部に通知する手段と、

前記送出要求情報に基づいて出力回線を決定するためにスケジューリングを実行するスケジューラを有するスケジューリング処理部と、

前記スケジューラで決定された出力回線の帯域制御を行う入力バッファと、

前記スケジューリング処理部において送出権が決定した出力回線に対して、その出力回線内の各サービス品質クラスの遅延および帯域制御を行い、所定の帯域内であればパケット形式のデータを前記バッファから読み出し、所定の帯域外の際には、前記バッファからパケット形式のデータを出力することなく、当該送出要求情報が無効である旨を前記スケジューラに通知する帯域制御部とからなる交換機。

【請求項 13】 送出要求情報とともに、サービス品質クラスをスケジューリング処理部に通知する手段と、

前記送出要求情報に基づいて出力回線を決定するためにスケジューリングを実行するスケジューラを有するスケジューリング処理部と、

前記スケジューラで決定された出力回線の帯域制御を行うバッファとからなり、

前記スケジューリング処理部は、前記サービス品質クラスを解析し、所定の帯域でかつ送出要求のある出力回線に対してスケジューリングを実行するパケット装置。

【請求項 14】 前記請求項 10, 11, 12 または 13 において、前記スケジューリング処理部は、帯域制御によっていずれのサービス品質クラスについても送出権が得られないときに、帯域外の任意のサービス品質クラスでの送出要求に対して送出権を与えるスケジューリング制御装置。

【請求項 15】 送出要求情報とともに、サービス品質クラスをスケジューリング処理部に通知する手段と、

前記送出要求情報に基づいて出力回線を決定するためにスケジューリングを実行するスケジューラを有するスケジューリング処理部と、

前記スケジューラで決定された出力回線の帯域制御を行うとともに、論理出方路に分割されたバッファと、

前記バッファにおいて、回線またはサービス品質クラス毎に設けられ、前記スケジューラの選択した出力回線に対しリーキバケットを計数する個別カウンタと、回線毎に設けられ、前回の選択からの経過時間を計数する代表カウンタにより帯域制御を行う帯域制御装置。

【請求項 16】 前記請求項 15 において、前記代表カウンタ部は内部タイマを備えており、各カウンタは前記内部タイマに基づいて前回の選択時における時刻を保持し、次の選択時に現時刻と比較することにより、経過時間を求める帯域制御装置。

【請求項 17】 m 個の入力方路からの入力セルをソーティングして、 m 個の出力方路に転送するソータを単位ソータとして、当該単位ソータを多段接続してなり、

第 1 段の単位ソータ群の単位ソータあたりの出力を、第 2 段に設けられたダミーソータと、第 2 段の単位ソータとに振り分けて出力し、

前記第 2 段のダミーソータと前記第 2 段の単位ソータとの出力を第 3 段の単位ソータ群に入力する、このようにして、 $(2N/m - 1) \times N/m$ 個の単位ソータを接続して全体として N 入力・ N 出力のソータ網を構成したパケット交換機。

【請求項 18】 m 個の入力方路からの入力セルをソーティングして、 N 個の出力方路に転送するソータを単位ソータとして、当該単位ソータを XY 方向にマトリクス状に多段接続してなり、

各単位ソータの出力を当該単位ソータの $+X$ かつ $-Y$ 方向に位置する次段の単位ソータに入力するとともに、

当該ソータ群の拡張時には、前記各単位ソータの出力を当該単位ソータの $+X$ 方向に位置する次段の単位ソータとする、このようにして、 $N/m \times (N/m + 1)$

）／2個の単位ソータを接続して全体としてN入力・N出力のソータ網を構成したパケット交換機。

【請求項19】 所望の出力回線に対するスケジューリング対象の各入力回線の送出要求情報を保持する要求情報管理部と、

スケジューリング開始入力回線を示すハイウェイ間ポインタを保持するハイウェイ間ポインタ制御部と、

各入力回線に対応した前記送出要求情報において、検索開始出力回線を示すハイウェイ内ポインタを保持するハイウェイ内ポインタ制御部と、

所望の出力回線に対する送出要求情報を保持する要求管理制御部と、

複数の前記送出要求情報の中からハイウェイ内ポインタ制御部が示す出力回線から検索を開始し、他の入力回線に選択されていない出力回線を選択するスケジューリング処理部とからなり、

前記スケジューリング処理部は、各入力回線からの有効セルに対する出力回線番号の割り当てを行った後、前記有効セルが割り当てられていない入力回線に対して空きセルに対するダミーの出力回線番号の割り当てを行い、全入力回線からのセルの出力回線番号を重複なく異なった値で転送するパケットスイッチ。

【請求項20】 所望の出力回線に対するスケジューリング対象の各入力回線の送出要求情報を保持する要求情報管理部と、

各出力回線間で選択優先度が異なるN（Nは2以上の自然数）個の優先度パターンを有し、該優先度パターンおよび該送出要求情報にしたがって他の入力回線が使用していない出力回線を選択するスケジューリング処理部と、

該優先度パターンの開始番号を示す優先度ポインタ制御部とを備え、

該スケジューリング処理部は、該優先度ポインタが示す該優先度パターンからNパターン分順次スケジューリングを行い、次のスケジューリング周期に該優先度パターンの開始番号を更新する際に、各入力回線からの有効セルに対する出力回線番号の割り当てと、空きセルに対するダミーの出力回線番号の割り当てとを競合制御することにより同時に行い、全入力回線からのセルの出力回線番号を重複なく異なった値で転送するパケットスイッチ。

【請求項21】 スケジューリング開始入力回線を示すハイウェイ間ポイン

タをスケジューリング周期毎に隣接する次の入力回線に更新し、入力回線に対応した送出要求情報に基づいて検索開始出力回線を示すハイウェイ内ポインタを送出が決定した回線に隣接する回線に更新するスケジューリング処理において、

前記各入力回線より転送する位相同期用のセルにシーケンス番号を付与し、

出力回線番号を順次インクリメントし、

各入力回線間の同一タイミングでの出力回線番号をシフトし、

スイッチ側の各格子点において、位相同期用のセルの受信タイミングとシーケンス番号とを比較することによってセルレベルでの位相同期を行うパケットスイッチング方法。

【請求項 22】 スケジューリング開始入力回線を示すハイウェイ間ポインタをスケジューリング周期毎に隣接する次の入力回線に更新し、入力回線に対応した送出要求情報に基づいて検索開始出力回線を示すハイウェイ内ポインタを送出が決定した回線に隣接する回線に更新するスケジューリング処理において、

前記各入力回線よりシーケンス番号を付与したセルをスイッチ内の各格子点に対してブロードキャスト転送し、

前記スイッチ内の各格子点において、到着したセルのシーケンス番号を比較することによりセルレベルでの位相同期を行うパケットスイッチング方法。

【請求項 23】 スケジューリング開始入力回線を示すハイウェイ間ポインタをスケジューリング周期毎に隣接する次の入力回線に更新し、入力回線に対応した送出要求情報に基づいて検索開始出力回線を示すハイウェイ内ポインタを送出が決定した回線に隣接する回線に更新するスケジューリング処理において、

前記各入力回線よりセルをスイッチ内の各格子点に対して転送し、

前記各格子点に設けられた位相調整用バッファでセルレベルでの位相差を調整するパケットスイッチング方法。

【発明の詳細な説明】

【0001】

【発明が属する技術分野】

本発明はスケジューリング制御装置に関し、特に大規模 ATM スwitch の一構成法である入力バッファ型 Switch におけるスケジューリング制御装置に適用し

て有効な技術である。

【0002】

【従来の技術】

近年、インターネットの爆発的な普及や大容量・高品質な情報を扱うメディアの登場により、大容量データを柔軟に扱うことのできる大規模な通信インフラの整備に期待が寄せられている。そして、実現の鍵となる数百ギガ～数テラオーダの容量をもつ交換機に関心が集まっている。

【0003】

入力方路毎に単一バッファをもつ基本的な入力バッファ型スイッチ（図2の左側）は、HOL（Head Of Line）ブロッキングの問題があり、スループットが約58.6%までにしか上がらないことが知られている。

【0004】

HOLブロッキングを回避する手段として、図3に示すように、入力バッファ部を論理的に出力方路毎に分割し、所定のアルゴリズムにしたがって送出権をスケジューリングする方式がいくつか提案されている。その一つとして入力と出力間でRequest/Acknowledge（要求・確認）制御を行う方式が知られているが、この方式は、入力ー出力間で何度も情報のやりとりを行って、スケジューリング処理を行っていた。また、入力と出力間の組み合わせが最大となるような組み合わせを求める別の方式では、最適な組み合わせを求めるために複雑な演算処理を繰り返し行っていた。

【0005】

入力バッファ部は出力回線分の論理バッファをもつが、サービス品質（QoS：Quality Of Service）での帯域制御を行うために、従来はその出力回線毎にQoS帯域制御用のリーキバケットカウンタ（LBカウンタ）が必要になった。例えば出力回線が128回線でQoSクラスが16クラスの場合、帯域制御用LBカウンタ数は2048個にもなる。

【0006】

従来、大規模入力バッファ型スイッチ向けスイッチ部の構成方法として、単位スイッチをクロスバー形式にする方式（図4）、バッチャソータ方式（図5）、

バッチャバンヤン方式（図 6）などがある。

【0007】

また、セルの同期方式としては、図 7 のようなセルの先頭を示すフレームパルスとデータを並送させることにより、セルの先頭の同期をとる方式がある。

【0008】

【発明が解決しようとする課題】

上述の Request/Acknowledge 制御を用いた方式では、特性向上のための入出力間で情報の受け渡しを繰り返し行う必要があることから、1 パケット時間内に処理を完了させるために高速なデバイスが必要であった。また、最大の組み合わせを求める様な方式は、複雑な論理演算が必要となり、これをハードウェアで実現することは困難であった。

【0009】

さらに、スケジューリング処理において、出力回線の決定に単純なラウンドロビン制御を用いる方式では、ラウンドロビン制御によって各回線が常に均等に割り当てられてしまい、一つの出力回線に対して、異なる入力負荷（不均等負荷）で複数回線から送出要求があった場合に、高負荷の入力回線からの送出遅延が著しく低下する問題があった。

【0010】

また、多数回線を収容する大規模スイッチにおいて、ロジックで帯域制御用の LB カウンタを構成するとハードウェア量が膨大になってしまう。また、これをメモリを使って構成すると 1 パケット時間に同時にアクセスすることが難しくなるという問題があった。

【0011】

また、複数の LSI、またはパッケージで大規模スイッチを構成する際、クロスバーだと LSI、またはパッケージ間の I/O 数がセルハイウェイの 4 倍分だけ必要となり、I/O（Input/Output）数がボトルネックとなる。

【0012】

また、バッチャソータ方式では I/O 数はセルハイウェイの 2 倍分で足りるが

、各ブロック毎の構成が異なり、またI/Oの長さも異なるため複数品種のLSI、またはパッケージを構成する必要が生じたり、位相調整が必要となるという問題もあった。

【0013】

また、複数LSI、またはパッケージで大規模スイッチを構成する際、バッチャソータ方式等では、スイッチ規模を拡張していく際に、最小規模時のパスに単位スイッチを挿入していく必要があるため、パスの張り替えが必要となり拡張時にシステムを停止させなければならないという問題があった。

【0014】

また複数LSI、またはパッケージで大規模スイッチを構成する際、スイッチ部を単位ソータのみで構成するには、全入力回線から重複なく異なる出回線番号をもつセルを常にスイッチ部に転送する必要があることから、送るべきセルがないとき、空きセルに対しても入力バッファ側で出回線番号をダミーで付与する必要がある。

【0015】

この点について、空きセルに対して入力バッファ側で出回線番号を付与することが、特開平3-36841号公報で提案されているが、この技術では、出力回線番号順に対応する入力回線を決定している。そのため、入力バッファを出方路毎の論理キューで構成し、競合制御により各入力バッファに対する出方路を決定するような出回線番号順に一意に入力回線を決定できないスケジューリング方式には適用できないという問題があった。

【0016】

またスイッチや入力バッファやスケジューラが複数のパッケージにまたがって構成されている場合、各パッケージ間でのセルレベルの同期外れが発生するが、図7に示したようなセルの先頭を示すフレームとデータを並送させる方式では、セルレベルの非同期は検出できないという問題があった。

【0017】

本発明は、これらの点に鑑みてなされたものであり、均等、不均等負荷のもとでも特性の劣化がなく、かつ、高速な繰り返しスケジューリングや複雑な演算処

理を必要とせず、簡素でかつ処理速度がデバイス能力に依存しないスケジューリング装置の実現を目的としている。

【0018】

【課題を解決するための手段】

図1および図8に本発明の基本的な構成を示す。

スケジューラ装置は、入力バッファ部からの送出要求に基づき、適切な出力方路を決定するものであり、各入力バッファから送出されるパケットを各々異なる出力方路にルーチングされるようにスケジューリングするものである。

【0019】

図8は本発明の原理構成図を示したものである。

要求情報管理部は、各入力回線が所望する出力回線への送出要求を入力回線毎に管理するものであり、送出要求数を管理するとともに、送出要求の有無を“0”，“1”で表す。

【0020】

負荷観測部は、一定周期内に到着した要求情報の数を、論理出力キュー毎に計測し、負荷を計測するものである。

ハイウェイ間ポインタ（PHW）は、ハイウェイ間ポインタ制御部に保持されており、スケジューリングの際に、どの入力回線からスケジューリングを開始するかを示す。このハイウェイ間ポインタ（PHW）が示す入力回線から順にN入力回線分のスケジューリングを行う。

【0021】

ハイウェイ内ポインタ（PO#j：jは回線番号）は、ハイウェイ内ポインタ制御部に保持され、所望する複数の出力回線の中から一つを選択する際に、どの出力回線から検索を行うかを示すものである。スケジューリング処理部はこのハイウェイ内ポインタ（PO#j）が示す出力回線から順に検索を行い、最初に見つかった回線に対して送出権を与えるようになっている。

【0022】

スケジューリング処理部は、ハイウェイ内ポインタとハイウェイ間ポインタが示す回線から検索を開始し、負荷観測部で得られた負荷に応じて、スケジューリ

ングを行う。

【0023】

スケジューリングは、送出要求のある回線で、ポインタ制御によって最初に見つかった回線（但し、他の回線に選択されていないこと）に対して送出権を与える。そして、送出権を与えた回線を次のスケジューリング処理に低優先になるようにポインタを更新するようになっている。

【0024】

ポインタの更新は、以下の手順で行われる。

図1に本発明の特徴的な一つのポインタ制御の動作を示す。ハイウェイ間ポインタは、全回線分のスケジューリングが終わった時点で、隣接する回線（右隣）に更新される。このとき、同一方向（時計回り）へN回更新された場合は、次のN回のスケジューリング処理は逆方向（反時計回り）の隣接回線へ更新するようになっている。

【0025】

ハイウェイ内ポインタは、スケジューリング処理部が決定した回線の次の回線にハイウェイ内ポインタ値を更新する。但し、ハイウェイ内ポインタが示す回線に要求がありかつその回線が既に他の回線によってスケジューリングされている場合には更新を行わないようになっている。

【0026】

本発明（請求項1および2に対応）によれば、ハイウェイ間ポインタが示す回線からスケジューリングを開始し、各回線におけるスケジューリング処理は、他の回線が使用していない回線を選択することにより、無駄なスケジューリング処理を回避し、かつ効率のよい（空きの少ない）スケジューリングが可能になる。

【0027】

また、スケジューリングされる入力回線の順番がHWポインタの反転制御によって、スケジューリング周期毎に反転するため、2つの入力回線で見たとときに、スケジューリングが行われる順序が均等に与えられる。

【0028】

また、本発明（請求項3）によれば、一つの回線についてスケジューリング処

理に時間差をもたせたパイプラインを形成することにより、1 パケット内のスケジューリング処理数が最大2回で済み、高速デバイスの適用無しに実現することができる。またパイプライン処理毎にハイウェイ内ポインタとハイウェイ間ポインタを備えることにより、一つのパイプラインシーケンスの中で、ハイウェイ内ポインタとハイウェイ間ポインタが均等に回り、均等スケジューリングが可能になる。

【0029】

また、本発明（請求項3～7）では、一定周期の負荷観測を行い、次の周期のスケジューリング時にその負荷を反映させることによって、負荷に応じたスケジューリングが可能となり、不均等負荷時におけるスケジューリングについても均等でかつ効率的なスケジューリングが可能となる。

【0030】

また、本発明（請求項4）では、一定周期内の到着パケット数を用いて、次の周期の最大送出数を前周期の到着パケット数に制限することによって、負荷の少ない回線は少なく、負荷の高い回線は多くパケットが送出できるようになる。

【0031】

また、本発明（請求項5～6）では、一定周期の到着パケット数を用いて、次周期の最優先出力回線の割合を負荷に応じて決定しており、前周期の到着パケット数分は最優先で送出が可能となる。また、到着パケット数で送出許可数を制限していないため、前周期に到着したパケット数以上のパケットも送出可能となり、効率的なスケジューリングが可能となる。

【0032】

また、本発明（請求項7）では、負荷観測によって得られた到着パケット数情報を、一定周期毎にクリアするのではなく、一定周期内で到着パケット数分の送出が行われずに残ったパケット数情報を、次のスケジューリングに繰り越すことにより、より厳密な負荷に応じてスケジューリングが可能となる。

【0033】

また、本発明（請求項8）では、一定周期毎のキュー長を用いて、次周期の出力回線の割合を決定しており、負荷に応じたスケジューリングが可能となる。

また、本発明（請求項 9）では、一つの入力回線についてスケジューリングを行うブロックを分散的にスイッチの入力ポート（入力バッファ）に配備することによって、スケジューリングの分散処理が可能になる。

【0034】

このとき、隣接する複数のスケジューリングブロックをリング接続して一つのグループを形成している。このグループは外部のグループと、内部のリングを切り替えるためのセレクタを備えており、新たにグループが増設されたとき、セレクタが外部側へと切り替わるようになっている。また、自グループから送出するデータを遅延ブロックにおいて遅延させることによって、増設前と後とでパイプライン処理時のタイムテーブル上の自スロット位置が変わらない構成となっている。

【0035】

このようなブロックを縦方向に拡張することによって、各スケジューラ間の接続リンクを張り替えることなく、オンラインで増設が可能になる。

また、本発明（請求項 10～14）では、到着パケットの送出要求を QoS クラス毎に管理し、スケジューラにおいて QoS の帯域、遅延を考慮してスケジューリングを行う。あるいは、スケジューラで出力回線のみを決定し、入力バッファ側の帯域制御部で、その出力回線の中でどの QoS クラスのパケットを送出するかを帯域、遅延特性を考慮して決定する。あるいは、遅延制御をスケジューラ側、帯域制御を入力バッファ側で行うなうことにより、パケットスイッチ全体および一つの回線内の帯域、遅延保証が可能となる。

【0036】

また、本発明（請求項 15, 16）によれば、回線毎の代表カウンタと、回線・QoS 個別の LB カウンタを設け、LB カウンタは、スケジューラにより選択された回線のみが使用され、代表カウンタは、前回選択されてからの経過時間を回線毎にカウントさせる。そして、LB カウンタでの制御時に経過時間分のカウントアップをまとめて行うことで、多回線の QoS 制御が可能となる。このとき、従来技術の図 4 に示したようなクロスバー方式だと、LSI、またはパケット間の I/O 数がセルハイウェイの 4 倍分必要となり、I/O 数がボトルネック

となる。一方、バッチャソータ方式（図5）の場合、I/O数はセルハイウェイの2倍分で済むが、各ブロックの構成が異なり、またI/Oの長さも異なるため複数品種のLSI、またはパッケージを構成する必要が生じる。

【0037】

この点について、本発明（請求項17に対応）では、 $(2N/m-1) \times N/m$ 個の単位ソータを接続して全体としてN入力・N出力のソータ網を構成したため、大規模ソータを構成を一種類の単位ソータをLSI、またはパッケージを接続して構成できるため、各単位ソータのI/O数はセルハイウェイの2倍分で済む。さらに各単位ソータ間の長さは両端の部分を除けば、統一可能なため、線長遅延による遅延調整回路を削減できる。

【0038】

図9は本発明（請求項17に対応）の原理図を示している。同図において、1チップまたは1パッケージあたりのm本の入力回線より入力されるセルをセルのヘッダ領域に付与される出回線情報に基づいて降順、または昇順にソーティングして、m本の出力回線に出力できる単位ソータを N/m 個ならべて、N方路の入力バッファに接続する。さらに当該単位ソータのm本のハイウェイ2分割して、同図に示すように、次の段の単位ソータに接続する。このとき線長遅延調整用のダミーの単位ソータを除けば、この段は1段目より一つ分単位ソータが少ない。このような1段目と2段目を1組みとして、 N/m 組接続して、全体として、N入力、N出力のソータ網を構成する。

【0039】

図10に本発明（請求項18に対応）の原理図を示す。ここでは、単位ソータをXY方向にマトリクス状に多段接続する際に、各単位ソータの出力を、単位ソータの+Xかつ-Y方向に位置する次段の単位ソータに入力させるようにした。

【0040】

そして、ソータ群を拡張するときには、各単位ソータの出力を当該単位ソータの+X方向に位置する次段の単位ソータに入力させるようにした。

これはすなわち、1チップまたは1パッケージあたりのm本の入力回線より入力されるセルをセルのヘッダ領域に付与される出回線情報に基づいて降順、また

は昇順にソーティングして、 m 本の出力回線に出力できる単位ソータを N/m 個ならべて、 N 方路の入力バッファに接続したことになる。当該単位ソータの m 本のハイウェイ2分割して、半分を運用用、残りの半分を拡張用として割り当て、図10のように接続する。

【0041】

従来のバッチャースータ方式や前記請求項17の構成の場合、スイッチ規模を拡張していく際に、パスの張り替えが必要となるが、この点について、本発明（請求項18に対応）では、スイッチ規模拡張時に、拡張前のパスは変更する必要がないため拡張性に優れている。

【0042】

さらに、クロスバー方式（図4）では $N/m \times N/m$ 個の単位ソータを必要とするのに対し、本発明（請求項18）では、単位ソータ数が $N/m \times (N/m + 1)/2$ で実現可能であり、 $N/m \times (N/m - 1)/2$ 個分だけ単位ソータ数を削減することができる。

【0043】

図11に本発明（請求項19）の原理図を示す。前述したスケジューリングにより、有効セルに対して、各入力回線より転送するセルの出回線番号を決定した後、出回線番号が決定しなかった各入力回線に対して、有効セル確定済出回線番号を除いて順次若番号より、空きセルに対する出回線番号を割り当て、あらかじめ出回線番号を割り当てた有効セルとともにソータ網へ転送する。

【0044】

図12に本発明（請求項20）の原理図を示す。全入力回線、出力回線に対して同一入力回線、または同一出力回線での優先度は異なるように優先度をあらかじめ割り当てておき、その優先度にしたがって順次各入力回線に対する出回線番号を決定するスケジューリングにおいて、最優先である優先度1のスケジューリング時、キューにデータがある入力回線は有効セルに対する読み出し回線として確定するが、キューにデータのない入力回線に対しても空きセルに対する仮の確定回線とする。

【0045】

次に優先度 2 に対するスケジューリングを行う際、優先度 2 に対応するキューにデータがあり、既に有効セルに対して同一入力回線または同一出力回線が確定していた場合には、優先度 2 の有効セルの確定は行わないが、空きセルに対して同一入力回線または同一出力回線が確定していた場合には優先度 2 の有効セルの確定を行う。

【0046】

このとき優先度 2 と同一入力回線または同一出力回線の優先度 2 について空きセルに対する仮の確定が設定されている場合には仮確定をはずす。

その後、優先度 2 に対応するキューにデータがないものについて、同一入力回線、または同一出力回線について、有効セルに対する確定、または空きセルに対する仮の確定がないときは、当該優先度 2 を仮確定とする。

【0047】

以下同様の処理を全優先度について行い、最終的に有効セルに対する確定、または空きセルに対する仮の確定が行われているものを最終的に確定とする。

このように、本発明（請求項 19，20 に対応）によれば、ソータ網に入力されるセルの出回線番号は、重複なく、かつ全ての異なる値をとるため、ソータ部では単純にソーティングするのみでよく、図 6 に示したようなバンヤン網（Banyan）を追加する必要がなくなる。

【0048】

また、本発明（請求項 20 に対応）では、多少制御は複雑になるものの、有効セルと空きセルに対する出回線番号の割り当てを同時に実施できるためより高速化を図ることができる。

【0049】

図 13 および図 14 は本発明（請求項 21 に対応）の原理図である。システムの初期状態やスイッチの増設時等のセルレベルでの位相同期が必要となった際、スケジューラ側より各入力バッファに対して順次出方路番号、およびシーケンス番号が増加するような同期用のセルパターンを送出する。

【0050】

この際、隣り合う入力回線からのセルのシーケンス番号は一致させ、出方路番

号は1ずつシフトしておく。このような入力セルに対して、各格子点の単位スイッチにおいて、隣り合う入力回線より到着した同一出方路の同期用セルの位相差 $\Delta SN - \Delta T$ (ここで、 ΔSN = 入力回線 # K から到着したセルのシーケンス番号 - 入力回線 # K + 1 から到着したセルのシーケンス番号、 ΔT = 入力回線 # K からのセルの到着時間 - 入力回線 # K + 1 からのセルの到着時間) を計算し、入力回線 # K + 1 に対応する入力バッファに対して、観測した位相差を通知する。入力バッファ # K + 1 では、受信した位相差をもとにスイッチへのセル挿入タイミングをシフトさせることにより、入力回線 # K と入力回線 # K + 1 より到着するセルの同期をとる。

【0051】

本発明(請求項21に対応)によれば、パッケージ間で位相同期が可能なため、位相調整バッファ量を少なくすることができる。また、通常セルと位相同期用のセルの出回線番号のコーディングが同一であるため、スイッチ内のセルのルーティングに関しては通常セル処理を共用できる。

【0052】

図15は本発明(請求項22に対応)の原理図である。システムの初期状態やスイッチの増設時等のセルレベルでの位相同期が必要となった際、スケジューラ側より各入力バッファに対して通常セルでは使用しない出方路番号(例えば出方路番号の最大値+1)を付与することによりブロードキャストセルであることを示し、さらに各スケジューラより入力するセルにシーケンス番号を付与し、当該シーケンス番号は同一タイミングで挿入するセルに関しては全入力回線で一致させて順次インクリメントしながら送出する。ブロードキャストセルを受信した各単位スイッチでは、ブロードキャストセルに対しては、全ての出方路へ転送するとともに、隣り合う各入力回線より転送されるシーケンス番号の差 ΔSN を計算し、 $\Delta SN \neq 0$ のときは、前述の本発明(請求項21)と同様に位相差 ΔSN を入力バッファ側へ通知し、受信した位相差に基づいて、送出タイミングをシフトすることにより、到着するセルの同期をとる。

【0053】

このように、本発明(請求項22)では、パッケージ間で位相同期が可能なた

め、位相調整バッファ量を少なくすることができる。前記請求項 21 に比べて、ブロードキャストセルに対して、全出方路にセルを転送する機能を追加する必要があるものの、全単位スイッチに常にブロードキャストセルが流れるため、任意のタイミングでシーケンス番号を比較することができ、より高速に位相調整処理を実現できる。

【0054】

図 16 は本発明（請求項 23 に対応）の原理図である。各格子点に位相外れ吸収用のバッファを保持し、セルレベルでの位相外れを吸収するようになっている。

【0055】

このように、本発明（請求項 23）では、同期が不要なため、同期用の回路を削減でき、またシステムの立ち上げや増設時に同期処理に要する時間を削減できる。なお、各格子点にバッファを設けるという点ではクロスポイントバッファ側スイッチと同様の構成となるが、入力バッファ側で、各格子点より各出回線に転送するための帯域は確保されているため、パッケージ間の遅延吸収分のバッファのみでよいいためバッファ量は格段に少なく済む。

【0056】

【発明の実施の形態】

以下、図面に基づいて、本発明の実施の形態を説明する。

【0057】

【実施例 1】

図 17 は、本システムの全体構成を示している。

同図に示すように、本実施例では、入力回線から到着したセルは入力バッファに保持され、スケジューラからの指示により $N \times N$ 構成のスイッチに入力され、 N 本の出力回線より出力される。

【0058】

ここで、図 18 を用いて、ハイウェイ間ポインタの反転制御の説明を行う。図 18 では、 4×4 個構成のスイッチを想定している。

同図において、1回目のスケジューリングは、入力回線#0から順に入力回線#0、#1、#2、#3の順でスケジューリングが行われる。

【0059】

もし、スケジューリング順序が一方向のみの場合、2回目はハイウェイ間ポインタが順方向に移動して、入力回線#2から順に、#2、#3、#0、#1の順でスケジューリングが行われる。ここで、入力回線#0と入力回線#1に着目してスケジューリング順序について見た場合、若番の入力回線の方が多く先にスケジューリングされてしまい不均等なスケジューリングになる（図中、入力回線#0と回線#1との間で先にスケジューリングされるものを四角で囲った）。

【0060】

一方、これに対して本実施例では、反転制御を用いる場合の2回目のスケジューリングにおいて、反転制御によって、HWポインタが逆方向に働き、入力回線#3から順に#3、#2、#1、#0の順でスケジューリングが行われる。

【0061】

このように、スケジューリング周期毎に反転制御を行うことで、スケジューリング周期毎にスケジューリング順序が入れ替わり、2周期のスパンで見たときに、均等なスケジューリングを実現できる。

【0062】

【実施例2】

以下に上記反転制御を用いたパイプライン処理系の実施例を示す。

上記実施例1では、4×4のスイッチ構成を例にあげており、1パケット時間内に4つの入力回線についてスケジューリングを行えばよいが、多くの回線を収容する場合には、1パケット時間内に全ての回線の処理を終えることが難しくなる。

【0063】

そこで、スケジューリング開始入力回線を時間的にずらしてパイプライン処理構成を形成することにより、処理速度面の課題を回避するのが本実施例である。

図19にパイプラインシーケンスを示す。各パイプラインの開始番号は時間的にずれており、競合が起こりにくくなっている。図中の数字は、スケジューリン

グ対象となる入力回線番号を示しており、この例では1パケット時間中に2つの回線のスケジューリングを実行する。そして、1回のスケジューリング周期が8パケット時間で完了するように組まれている。

【0064】

パイプライン#1では、Time1～Time8までが1周期のスケジューリングとなっており、上記実施例1で述べたように、Time9～Time16が反転周期のスケジューリング周期となっている（白抜き部分が順方向の処理、黒塗り部分が逆方向の処理）。

【0065】

パイプライン#2も同様であるが、スケジューリング開始回線が入力回線#2になっている（Time=2）。これは、パイプライン処理#1の開始番号から2個シフトした値である。

【0066】

実施例ではシフト量を2にしているが1個シフトでもよい。実施例では2個シフトする代わりに、パイプライン処理4以降で、1パケット時間内の処理順序を反転させている。

【0067】

パイプライン処理において、公平なスケジューリングを行うため、各パイプライン処理毎にハイウェイ間ポインタとハイウェイ内ポインタとをそれぞれ備えている。ここで本実施例のハードウェア構成は、図20および図21に示す通りとなる。図20はスケジューリング部を集中的に管理する場合のイメージであるが、図21のように各入力回線毎に分散的に配置することもできる。

【0068】

次に、ハイウェイ間ポインタの動作について説明する。

例えばTime=1のパイプライン処理#6における入力回線#0のスケジューリングにおいて、出力回線が回線#2に決定されたとする。このとき、次の入力回線#0のスケジューリング時に出力回線#2が低優先で送出されるようにパイプライン処理#6のハイウェイ内ポインタ（P HW#0__pipe#6）を3に更新する。

【0069】

Time = 3において、全体で見た場合の入力回線#0のスケジューリングが存在するが、このpipeline #4と前述のpipeline #6は完全に独立に動作しており、Time = 3ではpipeline #4のハイウェイ内ポインタにしたがってスケジューリングされる。

【0070】

ここで、pipeline #6に着目してみると、Time = 11において、次の入力回線#0のスケジューリング処理が存在する。これは、Time = 1で更新されたハイウェイ内ポインタ値(=3)にしたがって、出力回線#3から検索が行われる。

【0071】

このように時系列で見た場合には、各入力回線内の送出ポインタが均等に廻っていないようにみえるが、一つのパイプライン処理の中では均等にポインタが廻っており、全体として、均等な読み出しスケジューリングが実行されている。

【0072】

パイプライン処理におけるタイムテーブルの順序は、上記実施例の順序に限らず、どのようなパターンでもよく、パイプライン処理毎に一つずつ開始回線番号をずらすようにしてもよい。また、1パケット時間に複数の入力回線のスケジューリングを行ってもよいし、一つのスケジューリング周期の中で1パケット時間おきにスケジューリングを実行しないようにしてもよい。

【0073】

【実施例3】

次に、入力負荷に応じたスケジューリング方式の実施例を示す。

図22に不均等負荷時の例を示す。入力#0から出力#0に対する入力トラヒックが0.1の負荷で入力され、入力#1から出力#0に対する入力トラヒックが0.9の負荷で入力されているとする。

【0074】

この例に対して、ラウンドロビン制御を用いてスケジューリングを行った場合、入力#0と#1の選択される確率(すなわち読み出し確率)は各々0.5とな

り、パス # 1 - # 0 は入力 が 0.9 であることから、出力 負荷 に対して 入力 負荷 が 高い ため、キューイングされるパケットが増加し、キューイング遅延が大きくなってしまう。

【0075】

このような不均等負荷の場合にも、負荷に対して均等にスケジューリングする様に、負荷観測結果に基づいてスケジューリングを行う実施例を、図 23 を用いて説明する。

【0076】

スケジューリング装置は、一定周期毎に各論理出力キューに対する到着パケット数を計測する。ここでは、例として負荷観測周期を 10 パケット時間とし、到着パケット数が各々以下であるとする。

【0077】

入力 # 0 に到着し出力 # 0 に送出要求のあるパケット：1 個（負荷 0.1）

入力 # 1 に到着し出力 # 0 に送出要求のあるパケット：9 個（負荷 0.9）

次の周期（一定周期）では、上記到着パケット数に基づいてスケジューリングを行う。

【0078】

一定周期に送出可能なパケット数を前周期の到着パケット数に制限することにより、不均等負荷の場合でも高負荷パスの遅延は大きくなることはない。この例では、10 パケット時間の間にパス # 0 - # 0 のパケットは、いくら送出要求数が多くても、1 パケットしか出すことができない。そして、パス # 1 - # 0 は 9 個までは送出することができるため、不均等負荷時にも高負荷パスのキューイング遅延が大きくなることはない。

【0079】

すなわち、負荷の少ない回線に対して、大量に読み出されないように規制を行っている。

実際の動作例を、図 24 のフローチャートを用いて説明する。

【0080】

パケット（あるいは送出要求情報）が到着したとき、対応するパスの負荷観測

カウンタをインクリメントし、一定周期毎に、負荷観測カウンタの値をワーク用の到着パケット数ワークカウンタに加算する。これは、競合制御等によって、前周期に到着したパケット数分のパケットが全て送出しきれなかった場合に、残りの到着パケット数を次周期のスケジューリングで読み出すための処理である（図 24（a））。

【0081】

次に図 24（b）を用いて、スケジューリング時の動作について説明する。

スケジューリング時は、ハイウェイ内ポインタが示す出力回線の到着パケット数ワークカウンタをチェックし、1以上で、かつ、送出要求があればその回線に送出権を与える。このとき、当然ながら、他の入力回線に確保されていないことが必要である。そして、到着パケット数ワークカウンタをデクリメント（-1）する。

【0082】

一方、到着パケット数ワークカウンタが0であれば、送出要求の有無に関わらず、送出要求はないものと見なし、他の出力回線の送出が可能かをチェックする。

【0083】

このように、スケジューリング時は、この到着パケット数ワークカウンタを用いて、入力負荷以上とならないように送出パケット数に制限を設けている。

【0084】

【実施例 4】

実施例 3 では、到着パケット数に応じて送出パケット数を制限することによって、不均等負荷時の均等読み出しを実現したが、送出パケット数に制限をかけずに、出力回線の割合を到着パケット数にしたがわせてもよい。以下に動作例を示す。

【0085】

到着パケット数の計測は実施例 3 と同様であるが、到着パケット数が、次周期の最優先選択回線の割合を決定するために用いられる点で異なる。

例えば、上記実施例と同様の到着パケット数があったとする。

【0086】

入力#0に到着し出力#0に送出要求のあるパケット：1個（負荷0.1）

入力#1に到着し出力#0に送出要求のあるパケット：9個（負荷0.9）

この場合、次周期の最優先出力回線は、パス#0-#0：1回、パス#1-#0：9回となる。

【0087】

最優先出力回線は、到着パケット数に応じて決定され、最優先ポインタによって示される。最優先ポインタは、ハイウェイ内ポインタよりも上位のレベルに位置づけられ、スケジューラは、スケジューリングを行う際に、まず最優先送出ポインタをチェックする。

【0088】

このとき、最優先ポインタが示す回線に送出要求があれば、その回線を出力回線と決定（但し、送出要求があり他の回線に使用されていないこと）し、なければ、ハイウェイ内ポインタにしたがって、通常通りのスケジューリングを行い出力回線を決定する。

【0089】

次に、図25に実施例3との違いを示す例を示す。

・前周期の到着パケット数

：パス入力#0-出力#0：1個

：パス入力#0-出力#1：3個

・現周期の現時刻までの到着パケット数

：パス入力#0-出力#0：2個

：パス入力#0-出力#1：4個

実施例3の場合は、現周期に送出可能なパケットは、前周期の到着パケット数に制限されるため入力#0=1個、入力#1=3個までである。そしてこの場合では、たとえ現周期において空きがあっても、これらの数以上のパケットは送出することができず、現時刻までに到着した入力#0=2個、入力#1=4個のパケットは送出することができない。

【0090】

一方、本実施例は、現周期内で入力#0=1個、入力#1=3までは最優先で送出されるが、残りはHWポインタにしたがって均等に送出されるため、回線が空いていれば現周期に到着したパケットも送出することが可能である。

【0091】

図25の下線部の数（パス#0-#0：2個、パス#1-#0：4個）が空きのときに送出された数である。このときの現周期の到着パケット数は、パス#0-#0：2個、パス#1-#0：6個である。

【0092】

この場合、次周期の最優先で送出される数は、現周期に到着した数-前周期に最優先以外で送出された数となり、以下の式を適用するとパス#0-#0の場合、パス#1-#0の場合で以下になる。

【0093】

（現周期到着パケット数）-（前周期に最優先以外で送出された数）

=次周期の最優先数

パス#0-#0： $2 - 2 = 0$

パス#1-#0： $6 - 4 = 2$

最優先送出ポインタを用いたスケジューリングフローを図26に示す。

【0094】

第一に最優先ポインタ（パイプライン共通）が示す回線に送出要求があれば、送出確定とする。最優先ポインタが示す回線に要求がないか、あるいは到着パケット数ワークカウンタが0であれば、ハイウェイ内ポインタにしたがってスケジューリングを行う。

【0095】

出力回線が決定した後、出力回線に対応する到着パケット数ワークカウンタをデクリメントし、ポインタの更新を行う（図26中の右側フロー図）。

最優先ポインタの更新： 最優先ポインタが示す回線が送出確定になったとき、到着パケットワークカウンタをデクリメント（-1）し、最優先ポインタを現最優先ポインタの次に見つかる回線で、到着パケット数ワークカウンタが1以上

の回線に更新する。

【0096】

例えば、最優先ポインタが示す回線が #1 で送出確定回線も #1 であり、到着パケットワークカウンタが以下の通りであった場合は、次に見つかるカウンタ値 0 以上の回線は出力回線 #4 であることから、最優先ポインタは 4 に更新される。

【0097】

出力回線 #0 ワークカウンタ	: 4
出力回線 #1 ワークカウンタ	: 3
出力回線 #2 ワークカウンタ	: 0
出力回線 #3 ワークカウンタ	: 0
出力回線 #4 ワークカウンタ	: 5
出力回線 #5 ワークカウンタ	: 3

なお、ハイウェイ内ポインタは通常通り出力回線の次の回線、すなわち回線 #2 に更新される。

【0098】

最優先ポインタの動作を整理すると、ワークカウンタの値がそれぞれ以下の様な値であった場合、スケジューリング周期毎に最優先ポインタは下図のように変化する。

【0099】

入力回線 #3 ワークカウンタ	: 4 個
入力回線 #5 ワークカウンタ	: 2 個
最優先ポインタ値:	#3 (1回目) → #5 (2回目) → #3 (3回目)
	→ #5 (4回目) → #3 (5回目) → #3 (6回目)
	<u>→ #3 (7回目) → #3 (8回目) → #3 (9回目)</u>

上記において、下線部は最優先ポインタが #3 になっているが、ワークカウンタが 0 であるため、通常のハイウェイ内ポインタにしたがってスケジューリングされる。

【0100】

【実施例5】

上記実施例では、一定周期内に到着したセル数をベースとして、次の周期のスケジューリングを行っているが、一定周期毎に、入力バッファに蓄積されているパケット量、すなわちキュー長をもって、次の周期のスケジューリングを行ってもよい。図27に処理例を示す。

【0101】

一定周期毎に入力バッファ部のキュー長をチェックし、そのキュー長に応じて、最優先出力回線を決定する方法を示す。入力回線#0-出力回線0のキュー長が10個、入力回線#1-出力回線#0のキュー長が5個であった場合、上記到着パケット数ベースの実施例における、到着パケット数ワークカウンタにこの値をセットする（ここでは、到着パケット数ワークカウンタとしてではなく、キュー長ワークカウンタとして用いられる）。

【0102】

そして、最優先ポインタの決定時に本ワークカウンタをチェックし、キュー長ワークカウンタが0以上の回線を最優先ポインタにセットする。

図28(a)にワークカウンタ決定フローを、図29(a)にワークカウンタ設定例を示す。

【0103】

また別のやり方として、キュー長の割合から次周期のキュー長ワークカウンタを決定してもよい。フローチャートを図28(b)および(c)に示す。

全キュー長（同一出力方路のキュー長の和）に対して、自キュー長が占めている割合で次周期のワークカウンタ値を決定する方法を示す。

【0104】

図29(b)では、一定周期を20パケット時間としており、例えば、入力回線#0のワークカウンタ設定値は、自キュー長=10パケット、全キュー長=27であり、割合は、約0.37…となる。

【0105】

そして、次周期内における割合は、 $0.37 \times 20 = 7.4$ として7が求められる。小数点以下は、切り捨ててもよいし、切り上げてよい。同様に全入力回

線について、次周期のワークカウンタ値を決定する。

【0106】

また、自キューのキュー長の大小関係で、次周期のワークカウンタの値を決定してもよい。ここでは、キュー長0~3までを0.05の割合、3~10までを0.1の割合、10~20を0.2の割合を与えるとする。そして、一定周期を20パケット時間とする。

【0107】

そして、一定周期毎にキュー長をチェックし、現キュー長がどの範囲に入っているかを確認し、キュー長に応じた割合を得る。例えば、図29(c)の入力回線#0については、現キュー長=10であり、範囲3~10のため割合は、0.1が得られる。そして、次周期における割合は、 0.1×20 で得られ、2が求められる。

【0108】

上記実施例の範囲とその割合は、任意に設定できるものであり、実トラヒックに応じて決定してもよいし、あるトラヒック条件下のもと、適切に割合が振り分けられるように設定してもよい。

【0109】

キュー長が決定された後は、上記実施例4と同様に最優先ポイキューされる。フローチャートを図30に示す。動作は実施例4と全く同じである。

【0110】

【実施例6】

次にスケジューラの分散拡張方式について説明する。

これまで説明してきたスケジューリング装置は、図31のように各入力回線毎に分散して配置することができる。各入力回線毎のスケジューラは、自回線の要求情報管理部と、自回線のハイウェイ間ポイントと、パイプライン処理毎のハイウェイ内ポイントを有し、自回線のスケジューリングを行う。

【0111】

(図32)に示したような場合

）は、次にどの入力回線のスケジューリングを行うかを決定するために、ハイウェイ間ポインタが必要になるが、分散的に配備される場合、ハイウェイ間ポインタの値は常に自回線となり、ポインタとしては使用されない。その代わり、いつ自回線のスケジューリングを行うかの判定を、あらかじめ決められたタイムテーブルの自スロット位置で判断している。

【0112】

このような分散配備の場合には、スケジューリングの際に他の入力回線によって、送出要求のある出力回線が既に確定されているか否かを知るためにスケジューラ間で情報の受け渡しが必要となる。

【0113】

図32に分散配備時の接続構成例を示す。図32(a)のリング型接続は、各スケジューラを数珠つなぎ状に接続する。この場合、最下位のスケジューラと最上位のスケジューラ間の伝搬距離が長くなる。また、スケジューラ増設の度に、スケジューラ間の情報伝達リンクを張り替える必要がある。一方、図32(b)の様に、情報伝達リング状に均等にスケジューラが配置されるようにした場合、各スケジューラ間の伝搬遅延が等しくなり、またオンラインで張り替え無しに拡張することが可能となる。

【0114】

以下にスケジューラ接続の詳細を示す。

図33は、複数のスケジューラからなる拡張ユニットの構成と、その拡張方法を示したものである。

【0115】

図33は、4×4個構成のスイッチを8×8個構成から12×12個構成に拡張する方法を示している。

増設は、複数のスケジューラが集まった拡張ユニット毎に行われる。この例では、増設単位を4×4として、一つの拡張ユニットの中に4つのスケジューラを備える（図中では、2回線分のスケジューラを一つのブロックで表しているため、ブロックとしては2個であるが、論理的には4つである）。

【0116】

拡張ユニットは、これらのスケジューラをリング接続するとともに、他の拡張ユニットとの外部インタフェースを有し、外部インタフェースと内部リングとを切り替えるためのセレクタを有している。また内部リング上にスケジューラ間の情報伝達を遅らせる遅延ブロックを有する。例えば、図33の左側図では、拡張ユニットが一つのため、セレクタ部は、全て内部リング側を選択するようになっている。そして、スケジューリングを行う回線の順序は、回線#0→回線1→回線#2→回線#3で行われる。遅延制御ブロックは、回線1と回線#2の間にあり、回線#1のスケジューリング結果は、遅延ブロックにおいて一定時間遅らされて、回線#2のスケジューラに受け渡される。

【0117】

次に拡張ユニットを2個に拡張して、8×8のスイッチを構成する場合を示す。

拡張ユニットが増設された場合には、外部インタフェースにクロック、フレーム信号等が流れ始め、これらの信号の検出を行うことにより、各拡張ユニットのセレクタは、外部インタフェース側を選択するようになる。そして、元々存在していた拡張ユニットの遅延ブロックは使われなくなり、代わりに新しく増設された方の拡張ユニットの遅延ブロックが働く。スケジューリング順序は、回線#0→回線1→回線4→回線5→回線6→回線7→回線#2→回線#3の順で行われ、遅延ブロックは回線#5と回線#6の間で用いられる。

【0118】

12×12に増設する場合も同様の手順で行われる。このように、スケジューラ増設の際にオンラインで、かつスケジューラ間の情報リンクの張り替え無しに増設することが可能になる。

【0119】

次に、遅延ブロックの実際の動作について述べる。

図34に遅延ブロックの働きを示す。ここでは、最大スイッチ構成を8×8と仮定した場合を例にとってあげる。

【0120】

まず、パイプラインシーケンスは、最大構成時を想定して図34の下段図のよ

うに組まれたとする。各入力方路毎の分散スケジューラは、このパイプラインシーケンスの自回線番号のロットにおいてスケジューリング動作を行う。

【0121】

回線数の増加等によって、接続回線数が変わる度に自タイムスロットが変更されることは、オンラインで増設する際の切り替え時の処理が複雑になったり、ハードウェアの作りが難しくなる。そこで、最大構成時のスケジューリングシーケンスにあわせて、少ない回線時にも各スケジューラ間のデータ転送を遅らせることによって、自タイムスロットを常に固定化させている。

【0122】

図34の上段図は、4×4のときのパイプラインシーケンスである。スケジューリング順序は回線#0→回線#1→回線#2→回線#3であるが、回線#1のスケジューリング結果の情報を4パケット時間遅延させて回線#2に送信する。このようにすることによって、自タイムスロットの位置を固定化することができる。図中、パイプライン#1, 2, 3, 5, 6, 7はスケジューリング情報を遅延させていないが、これは、遅延ブロックの一つ前のスケジューラがスケジューリング周期の最後の回線であるため、後段の回線に情報通知が不要のためである。

【0123】

各パイプラインシーケンスと遅延との関係を以下に示す。

4×4	0 1 ----->	2 3 (遅延 8)
8×8	0 1 4 5 ----->	6 7 2 3 (遅延 4)
12×12	0 1 4 5 8 9 10 11 6 7 2 3	(遅延 0)

【0124】

【実施例7】

入力バッファ、スケジューラ両方でQoS制御を行う場合の一実施例を示す。

【0125】

図35に、入力バッファ部とスケジューラ部の関係を示す。入力バッファ部は、到着したパケットをN方路の論理出力キュー毎、かつmクラスのQoSクラス

毎に管理する。そして、スケジューラ部へ、送出要求情報（所望する出力方路番号）とQoSクラスを通知する。

【0126】

スケジューラの要求管理部は、論理出力キュー毎、かつQoSクラス毎に、その情報を管理する。

スケジューラ部では、QoSクラスの遅延特性のみを考慮し、高優先クラスの送出要求からスケジューリングを行い、高優先クラスに送出対象がなければ、次の高優先クラス送出が可能かを判定し、高優先クラスから順にスケジューリングを行う。

【0127】

実際の処理順序を以下に示す。

例えば、出力回線#1のQoSクラス#3と、出力回線#2のQoSクラス#0とに要求があったとする。そして、ハイウェイ内ポインタは#0を示しているとする（QoSクラス0が最高優先クラス）。

【0128】

まず、高優先クラスについてスケジューリングを行う。この例ではQoSクラス#0の要求があるのは、出力回線#2のみのため、仮の送出許可出力回線を#2と決定する。このとき、他の回線にも同じくQoSクラス#0の要求があった場合には、ハイウェイ間ポインタにしたがって、回線#1から順に検索し、最初に見つかった回線に仮の送出許可出力回線を与える。

【0129】

そして、仮の送出許可出力回線は、QoSクラスとともに入力バッファに通知される。

入力バッファ部では、受け取った情報をもとに、仮出力回線のQoSクラスの帯域制御を行い、所望の帯域で出力可能かをチェックし、帯域内であればバッファ部に対して読み出し指示を与える。ここで、仮に帯域外であったと判定された場合には、バッファからはパケットを送出せず、スケジューラ部へ帯域外であったことを通知する。

【0130】

帯域外である旨の通知を受け取ったスケジューラ部では、次の優先クラスのスケジューリングを行う。この例では、QoSクラス#1と#2の要求がないため、その次のQoSクラス#3のスケジューリングを行う。

【0131】

QoSクラス#3の要求をもっているのは、入力回線#1のみである。したが、再び入力バッファ部へ、仮の出力回線番号とQoSクラスを通知し上記処理を繰り返す。

【0132】

これら一連の処理によって、スケジューラ部で遅延特性、入力バッファ部で帯域特性を保証することが可能になる。

この動作フローを図36に示す。

【0133】

【実施例8】

上記実施例では、QoSクラスの最優先クラスから順にスケジューリングを行い、出力回線が確定するまで繰り返し処理を行っているが、下記に示す様な方法をとってもよい。

【0134】

入力バッファからは、送出要求情報とQoSクラスをスケジューラ部に通知し、スケジューラ部は、このQoSクラス毎の情報をK個のグループにマッピングする。

【0135】

例えば、6つのQoSクラスを、QoSクラス#0～4番をグループ#0、QoSクラス#5をグループ#1にマッピング。

要求管理部では、到着した要求情報を出力回線毎、かつグループ毎に管理する。

【0136】

そして、スケジューラはこのグループ中の最優先グループからスケジューリングを行い、上位のグループにおいて出力回線が確定しない場合に、その次の低位グループのスケジューリングを行う様に制御し、入力バッファ部へは、確定した

送出許可出力回線番号とグループ番号を通知する。

【0137】

入力バッファのQoS制御部では、そのグループ内のQoSクラスの中で、さらに最優先クラスのものから順に帯域制御を行い、帯域内であればバッファに読み出し指示を与える。

【0138】

この処理フローを図37に示す。

上記実施例では、完全に出力回線が確定した後に、入力バッファへ送出許可回線番号を返却しているが、グループ数が少ない場合には、実施例6のように仮の送出許可回線番号を一旦入力バッファ部へ通知して、帯域外のとき、次のグループのスケジューリングを行うようにしてもよい。

【0139】

また、グループ数やマッピングされるQoSの組み合わせは、上記実施例に限らず、1～Nのグループに任意にマッピングしてもよい。

【0140】

【実施例9】

次に、別方法のQoS制御の一実施例を示す。

図38に、入力バッファ部とスケジューラ部の関係を示す。入力バッファ部は、到着したパケットをN方路の論理出力キュー毎、かつmクラスのQoSクラス毎に管理する。そして、スケジューラ部へは、送出要求情報（所望する出力方路番号）のみを通知する。

【0141】

スケジューラの要求管理部は、論理出力キュー毎に、その情報を要求管理部で管理する。

スケジューラ部では、QoSクラスを考慮せず、要求出力回線にしたがってスケジューリングを行い、入力バッファに対して送出許可出力回線番号を通知する。

【0142】

送出許可回線を受け取った入力バッファ部では、その出力回線内のどのQoS

クラスの packets を送出するかを決定すべく、QoS 制御部において、遅延クラスの高いクラスで所定の帯域内のクラスを選択し、そのクラスの packets をバッファから送出する。

【0143】

ここで、どのクラスも所定の帯域外の場合には、バッファからは packets を送出せず、帯域外であったことをスケジューラ部に通知する。

帯域外であった場合には、スケジューラ部の要求管理部の要求情報数をデクリメントしないように制御する。

【0144】

本 QoS 制御方法の動作フローを図 39 に示す。

フローは、スケジューラ部がスケジューリングを開始する部分から出力回線を決定するまでを示している。

【0145】

【実施例 10】

次に、スケジューラ側で一括して QoS 制御を行う場合の実施例を示す。

図 40 に入力バッファとスケジューラの関係を示す。

【0146】

入力バッファからは、到着 packets の送出要求情報と QoS クラスをスケジューラに対して通知し、スケジューラ側で論理出力毎かつ QoS クラス毎に管理を行う。

【0147】

そして、スケジューラの QoS 制御部において、優先度の高い QoS クラスから順に帯域制御を行い、所定の帯域内の QoS クラスを選定する。そして、それらの情報をもとにスケジューリングを行い、入力バッファへは、確定した出力回線番号と QoS クラスを通知し、入力バッファ側では、受け取った情報にしたがってバッファから packets を送出する。

【0148】

このフローを図 41 に示す。

【0149】

【実施例 11】

上記実施例 7～10 では、帯域制御部において、帯域外の場合は送出要求があっても、送出することができない。そのため、全 QoS クラスの帯域制御において、全てのクラスが帯域外であった場合には、全くパケットを送出することができなくなってしまう。

【0150】

そこで、全 QoS クラスの送出が確定しなかった場合に、帯域外であるが、あらかじめ設定された QoS クラスに送出要求があればそのクラスに送出権を与えてもよい。

【0151】

送出クラスは、全クラスに与えられてもよいし、特定の QoS クラスにのみ与えられてもよい。また、設定クラスが複数あった場合には、スケジューリングする順序を最優先クラスから行ってもよいし、別途設けられた優先順序にしたがってもよい。

【0152】

このフローを図 42 に示す。

【0153】

【実施例 12】

図 43 に入力バッファ部の構成を、図 44 に QoS 制御部のブロック図を示す。

【0154】

QoS 制御部は、出力回線毎の代表カウンタと、出力回線・QoS 個別のリーキパケット (LB) カウンタを設ける。LB カウンタは、スケジューラにより選択された回線のみが使用され、代表カウンタは、前回選択されてからの経過時間を回線毎にカウントする。LB カウンタでの制御時に経過時間分のカウントアップをまとめて行うことで、多回線の QoS 制御においても処理数が少なくて済む。

【0155】

図 45 に、代表カウンタ部と演算部の構成例を示す。代表カウンタ部は、回線

毎の経過時間を保持するカウンタをもち、1パケット時間毎に全カウンタが1ずつカウントアップされる。スケジューラが読み出し回線を決定し、代表カウンタ部に通知されると、代表カウンタは決定した回線の代表カウンタ値を選択し、演算部へ通知する。同時に、その回線の代表カウンタを1にリセットする。これにより、回線毎の代表カウンタが、前回選択されてからの経過時間を測定することができる。

【0156】

演算部はQoS毎に用意される。まず、帯域情報のうち加算値に対し、通知された代表カウンタ値を掛け合わせ、それをLBカウンタ値に加えることで、経過時間分の加算をまとめて行う。その演算結果について、0以上の場合に読み出し候補としてQoS選択部に読出要求を出す。

【0157】

QoS選択部は、各QoSの演算部からの読み出し要求をもとに読み出しQoSを決定し、決定したQoSあるいは読み出し要求があるにも関わらずバッファが空であったQoSを、減算対象として演算部に通知する。減算指示のあったQoSの演算部はLBカウンタを減算し、その結果をLBカウンタ更新値としてLBカウンタ部に通知する。

【0158】

これにより、経過時間分のLBカウンタ処理を1パケット時間にまとめて行うことができる。本演算部のアルゴリズムを図46に示す。なお、本例では読み出し要求があり、バッファが空きであったQoSについては減算処理を行うが、これはシステムのポリシーにより減算を行わない、あるいはカウンタを止める等の処理を行ってもよい。

【0159】

ここで、経過時間カウンタがMAX値を越えてしまった場合に値が0に戻ってしまうと、十分時間がたっているのにも拘わらず加算値が小さくなり帯域外と判定される可能性がある。経過時間カウンタがMAX値を越えるということはその間入力パケットがないケースであると想定されるため、読み出し判定が遅れることも許容できるが、それを防ぎたい場合は、カウンタをMAX値で止めておくこ

とで対処可能である。

【0160】

次に、図47に代表カウンタの別の構成例を示す。ここではカウンタは用いず、前回選択した時間を保持しておくことで、経過時間を計測する。スケジューラより読み出し回線が通知されると、RAMより対応する回線の前回選択時間を読み出す。そしてその値と内部時計カウンタによって示される現在時刻の差分をとり、前回選択からの経過時間を求める。

【0161】

同時に、対応するRAMの値を現時刻に書き換える。これにより代表カウンタと同等の動作を行うことができる。RAMのアクセス数は1パケット時間にR/W一回ずつであり、実現可能である。よって、ハードウェアの大きいカウンタではなくRAMを用いて構成できるため、ハードウェア量が削減可能である。

【0162】

ここで、内部時計カウンタの周期を経過時間が越えてしまった場合、十分時間がたっているのにも拘わらず加算値が小さくなり帯域外と判定される可能性がある。経過時間が内部時計カウンタの周期を越えるということはその間入力パケットがないケースであると想定されるため、読み出し判定が遅れることも許容できる。しかしこれを防ぎたい場合は、経過時間保持RAMに有効フラグbitを追加し、有効フラグビットを周期監視することで対処可能である。

【0163】

図48に周期監視方式例を示す。周期監視用の回線指示カウンタを用意し、その回線指示カウンタに対応する経過時間に対し、有効フラグが有効である場合、現在時刻からの差分（経過時間）を算出する。そしてその差分が判定しきい値を越えていた場合、有効フラグを無効に書き換える。そして読み出し回線の経過時間算出時、経過時間が無効時は経過時間を内部時計カウンタのMAX値として通知する。これにより、経過時間が小さくなってしまうことを防ぐことが可能である。周期監視用RAMアクセスはR/W一回ずつの増加で済む。

【0164】

また、別の方法として内部時計の1周期をいくつかの領域に分割し、現在時刻

が属する領域にしたがって、クリア対象の領域に属する選択時間の有効フラグをクリアする方式がある。

【0165】

図27の左半図Aに、4つの領域に分割した場合の有効時間チェック方式の例を示す。例えば現在時刻がAに属する場合は、Bの領域について有効フラグのクリアを行う。領域の分割数は、一つの領域以内で全回線のチェックが終わるだけの時間以上あればよい。例えば分割数が4の場合、上位2bitだけをチェックすればよいので、回路が簡易化できる。

【0166】

また、入力パケットがしばらくなく経過時間が大きいケースにおいて、加算値後のLBカウンタが大きくなりすぎてしばらく読み出し対象になり続ける可能性がある。これに対し、LBカウンタに上限値を設け、計算結果が上限値を越えた場合は上限値に戻す処理を行うことで回避できる。

【0167】

【実施例13】

本実施例では、図9に示すように1チップまたは1パッケージあたりのm本の入力回線より入力されるセルをセルのヘッダ領域に付与される出回線情報に基づいて降順、または昇順にソーティングして、m本の出力回線に出力できる単位ソータを N/m 個ならべて、N方路の入力バッファに接続する。

【0168】

当該単位ソータのm本のハイウェイ2分割して、図9のように次の段の単位ソータに接続する。このとき線長遅延調整用のダミーの単位ソータを除けば、この段は1段目より一つ分単位ソータが少ない。このような1段目と2段目を1組みとして、 N/m 組み接続して、全体として、N入力N出力のソータ網を構成する。

【0169】

トータルのチップ数は線長遅延調整用のダミーの単位ソータを除けば、 $(2N/m - 1) \times N/m$ 個で構成される。

また、線長遅延調整用のダミーの単位ソータは線長遅延を別の回路で調整する

か線長遅延を単位ソータで吸収できるのならなくてもよい。

【0170】

【実施例 14】

本実施例では、図 10 に示すように、1 チップまたは 1 パッケージあたりの本数の入力回線より入力されるセルをセルのヘッダ領域に付与される出回線情報に基づいて降順、または昇順にソーティングして、 m 本の出力回線に出力できる単位ソータを N/m 個ならべて、 N 方路の入力バッファに接続する。当該単位ソータの m 本のハイウェイ 2 分割して、半分を運用用、残りの半分を拡張用として割り当て、図 10 のように接続する。図 10 の上段図と下段図とは等価構成を示している。

【0171】

また 1 単位ソータのみ使用する場合は単位ソータ A のみを使用し、拡張時は拡張用のハイウェイに単位ソータ B, C を接続し、順次単位ソータ E, D, F、単位ソータ I, G, H, J という順に接続しながらソータ網の規模を拡大する。

【0172】

【実施例 15】

本実施例では、図 11 および図 49 で示すフロー図のように前述のスケジューリングにより、有効セルに対して、各入力回線より転送するセルの出回線番号を決定した後、出回線番号が決定しなかった各入力回線に対して、有効セル確定済出回線番号を除いて順次若番号より、空きセルに対する出回線番号を割り当て、あらかじめ出回線番号を割り当てた有効セルとともにソータ網へ転送する。

【0173】

このとき、空きセルに対する出回線番号の割り当ては有効セルに対する番号割り当てとは異なり競合制御や回線間のフェアネスの考慮は不要なので、一つのラウンドロビンポインタを最若番号から最古番号へ一巡する間に決定できる。

【0174】

また、有効セルに対する出回線番号の割り当ては、スケジューリング処理部が、優先度ポインタを示す優先度パターンから N パターン分だけ順次スケジューリングを行い、次のスケジューリング周期に該優先度パターンの開始番号を更新

するようにしてもよい。

【0175】

【実施例 16】

本実施例では、図 12 および図 50 に示すフロー図のように全入力回線、出力回線に対して同一入力回線、または同一出力回線での優先度は異なるように優先度をあらかじめ割り当てておき、その優先度にしたがって、順次各入力回線に対する出回線番号を決定するスケジューリングを行うようにする。そして、最優先である優先度 1 のスケジューリング時、キューにデータがある入力回線は有効セルに対する読み出し回線として確定するが、キューにデータの無い入力回線に対しても空きセルに対する仮の確定回線とする。

【0176】

次に優先度 2 に対するスケジューリングを行う際に、優先度 2 に対応するキューにデータがあり、既に有効セルに対して同一入力回線または同一出力回線が確定していた場合には、優先度 2 の有効セルの確定は行わないが、空きセルに対して同一入力回線または同一出力回線が確定していた場合には優先度 2 の有効セルの確定を行う。

【0177】

このとき優先度 2 と同一入力回線または同一出力回線の優先度 1 について空きセルに対する仮の確定が設定されている場合には仮確定をはずす。

その後、優先度 2 に対応するキューにデータがないものについて、同一入力回線、または同一出力回線について、有効セルに対する確定、または空きセルに対する仮の確定がないときは、当該優先度 2 を仮確定とする。

【0178】

以下同様の処理を全優先度について行い、最終的に有効セルに対する確定、または空きセルに対する仮の確定が行われているものを最終的に確定とする。

【0179】

【実施例 17】

本実施例では、図 14 に示すようにシステムの初期状態やスイッチの増設時等のセルレベルでの位相同期が必要となった際、スケジューラ側より各入力バッフ

ァに対して順次出方路番号、およびシーケンス番号が増加するような同期用のセルパターンを送出する。

【0180】

この際に、隣り合う入力回線からのセルのシーケンス番号は一致させ、出方路番号は1ずつシフトしておく。単位スイッチの最小単位となるAより位相同期をはじめる場合、入力回線#1と入力回線#2より到着するセルの位相差として $\Delta SN - \Delta T$ を算出し、 $\Delta SN - \Delta T = 0$ のときはOKとし、 $\Delta SN - \Delta T \neq 0$ のときは、位相差を入力バッファ#2へ通知し、入力バッファ#2では受信した位相差に基づいて、セル送出位相を変更する。

【0181】

なお、 $\Delta SN - \Delta T = 0$ の判定処理は、 $\Delta SN - \Delta T = 0$ になるまで繰り返す。また、 $\Delta SN - \Delta T = 0$ となったとき、上段の単位スイッチBおよび右側の単位スイッチDに対して位相調整完了通知を行う。

【0182】

以上の処理を全単位スイッチの同期がとるまで繰り返す。なお、位相調整を行う単位スイッチが入力バッファに対して最近接でないときは、前段の単位スイッチと当該単位スイッチ間の位相調整用バッファについて、入力バッファと同様の処理を行うことにより実施する。

【0183】

図14は、あらかじめ各入力回線間の位相差が3, 2, 1のときの位相調整の動作例を示している。また同期の開始点は、拡張しない場合には、位相遅延の最も大きいG側から行ってもよい。

【0184】

なお、図13では各格子点が2入力2出力の最小の単位スイッチとなっているが、N入力N出力のスイッチを単位スイッチとして置き換えてもよい。

【0185】

【実施例18】

本実施例では、図15に示すようにシステムの初期状態やスイッチの増設時等のセルレベルでの位相同期が必要となった際、スケジューラ側より各入力バッ

ァに対して通常セルでは使用しない出方路番号（例えば出方路番号の最大値+1）を付与することによりブロードキャストセルであることを示し、さらに各スケジューラより入力するセルにシーケンス番号を付与し、当該シーケンス番号は同一タイミングで挿入するセルに関しては全入力回線で一致させて順次インクリメントしながら送出する。

【0186】

ブロードキャストセルを受信した各単位スイッチでは、ブロードキャストセルに対しては、全ての出方路へ転送する。単位スイッチの最小単位となるAより位相同期をはじめる場合、入力回線#1と入力回線#2より到着するセルの位相差として ΔSN を算出し、 $\Delta SN = 0$ のときはOKとし、 $\Delta SN \neq 0$ のときは、位相差を入力バッファ#2へ通知し、入力バッファ#2では受信した位相差に基づいて、セル送出位相を変更する。

【0187】

なお、 $\Delta SN = 0$ の判定処理は、 $\Delta SN = 0$ になるまで繰り返す。また、 $\Delta SN = 0$ となったとき、上段の単位スイッチBおよび右側の単位スイッチDに対して位相調整完了通知を行う。

【0188】

以上の処理を全単位スイッチの同期がとるまで繰り返す。なお、位相調整を行う単位スイッチが入力バッファに対して最近接でないときは、前段の単位スイッチと当該単位スイッチ間の位相調整用バッファについて、入力バッファと同様の処理を行うことにより実施する。

【0189】

また同期の開始点は、拡張しない場合には、位相遅延の最も大きいG側から行ってもよい。

なお、図16では各格子点が2入力2出力の最小の単位スイッチとなっているが、N入力N出力のスイッチと単位スイッチとして置き換えてもよい。またスイッチ構成としてはクロスバー型ではなくソータ網を適用してもよい。

【0190】

【実施例19】

本実施例では、図 16 に示すように各格子点に位相同期用のバッファを保持し、セルレベルでの位相外れを吸収する。

【0191】

なお、後段の単位スイッチのバッファと前段の単位スイッチの読み出し時の優先制御方法としては、単に後段のバッファにセルがあるときには、前段の単位スイッチのバッファからの読み出しを停止するバックプレッシャー型でもよいし、前段と後段につながる単位スイッチの数を考慮して、 $1/k : 1 - 1/k$ （但し k : 全回線数 - 回線番号）の割合で読み出すようにしてもよい。

【0192】

【発明の効果】

以上述べたように、本発明では、均等、不均等負荷のもとでも特性の劣化がなく、かつ、高速な繰り返しスケジューリングや複雑な演算処理を必要とせず、簡素でかつ処理速度がデバイス能力に依存しないスケジューリングシステムを実現できる。

【0193】

【図面の簡単な説明】

【図 1】 本発明の基本原理図

【図 2】 従来技術における入力バッファ型スイッチの問題点を説明するための図

【図 3】 従来技術における入力バッファ型スイッチの構成図

【図 4】 従来技術におけるクロスバー方式の説明図

【図 5】 従来技術におけるバッチャーソータ方式の説明図

【図 6】 従来技術におけるバンヤン網の説明図

【図 7】 従来技術におけるセルとフレームパルスとを同期させる技術の説明図

【図 8】 本発明のスケジューラ装置のブロック構成図

【図 9】 本発明（請求項 17 に対応）の原理図

【図 10】 本発明（請求項 18 に対応）の原理図

【図 11】 本発明（請求項 19 に対応）の原理図

【図 12】 本発明（請求項 20 に対応）の原理図

- 【図 13】 本発明（請求項 21 に対応）の原理図
- 【図 14】 本発明（請求項 21 に対応）の原理図
- 【図 15】 本発明（請求項 22 に対応）の原理図
- 【図 16】 本発明（請求項 23 に対応）の原理図
- 【図 17】 実施例 1 のシステム構成を示すブロック図
- 【図 18】 実施例 1 におけるハイウェイ間ポインタの反転制御を説明するための図
- 【図 19】 実施例 2 におけるパイプライン処理時のポインタ制御を説明するための図
- 【図 20】 実施例 2 におけるスケジューリング処理部を集中背部した場合のブロック図
- 【図 21】 実施例 2 におけるスケジューリング処理部を分散配備した場合のブロック図
- 【図 22】 実施例 3 における不均等負荷時における入力回線と出力回線との関係を示す説明図
- 【図 23】 実施例 3 における送出セル数を到着セル数に制限する場合のフロー図
- 【図 24】 実施例 3 における送出セル数を規制する場合のフロー図
- 【図 25】 実施例 4 における最優先送出回線を到着セル数の割合で決定する方法を説明する図
- 【図 26】 実施例 4 における最優先送出ポインタを用いたスケジューリングフロー図
- 【図 27】 実施例 5 においてキュー長を用いた最優先送出回線決定方法を説明するための図
- 【図 28】 実施例 5 におけるキュー長ベースの負荷観測のためのフロー図
- 【図 29】 実施例 5 においてキュー長ワークカウンタ設定値の決定方法についての説明図
- 【図 30】 実施例 5 におけるキュー長ベースの最優先送出ポインタを用いた制御フロー図

【図 3 1】 実施例 6 における分散スケジューラ型入力バッファスイッチを説明するための図

【図 3 2】 実施例 6 におけるスケジューラの拡張方法を説明するための図

【図 3 3】 実施例 6 における分散スケジューラ拡張構成を説明するための図

【図 3 4】 実施例 6 における遅延ブロックの動きを示す説明図

【図 3 5】 実施例 7 における QoS 制御のためのブロック図

【図 3 6】 実施例 7 における QoS 制御のフロー図

【図 3 7】 実施例 8 における QoS 制御のフロー図

【図 3 8】 実施例 9 における QoS 制御のためのブロック図

【図 3 9】 実施例 9 における QoS 制御のフロー図

【図 4 0】 実施例 10 における QoS 制御のためのブロック図

【図 4 1】 実施例 10 における帯域制御フロー図

【図 4 2】 実施例 11 における帯域制御フロー図

【図 4 3】 実施例 12 における入力バッファの構成図

【図 4 4】 実施例 12 における入力バッファの QoS 制御部のブロック構成図

【図 4 5】 実施例 12 における代表カウンタと演算部のブロック構成図

【図 4 6】 実施例 12 における演算部の動作アルゴリズムを示すフロー図

【図 4 7】 実施例 12 における代表カウンタの構成例を示す図

【図 4 8】 実施例 12 における選択時間の周期監視の説明図

【図 4 9】 実施例 15 における処理フロー図

【図 5 0】 実施例 16 における処理フロー図

【符号の説明】

QoS サービス品質

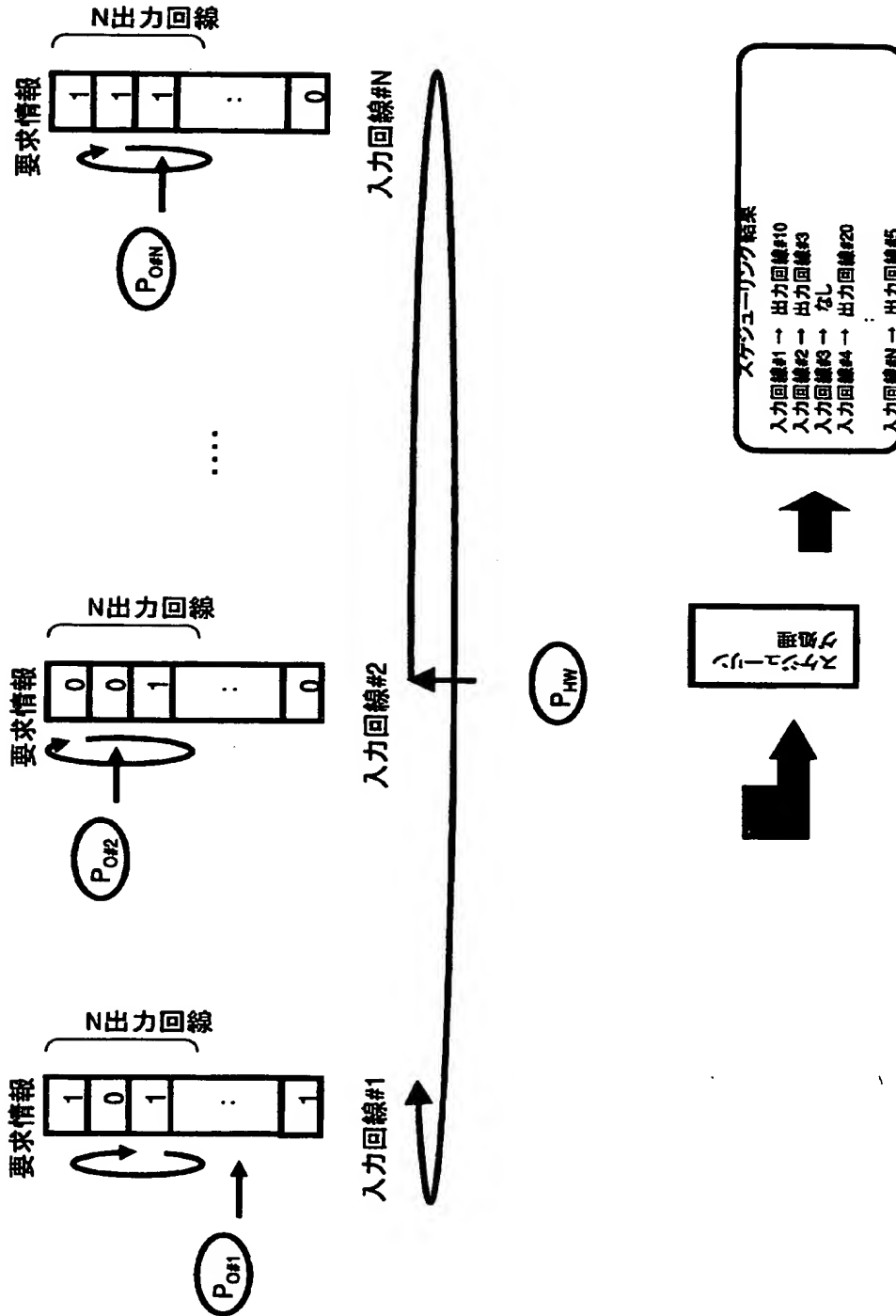
LB カウンタ リークバケットカウンタ

PHW ハイウェイ間ポインタ

PO#j ハイウェイ内ポインタ

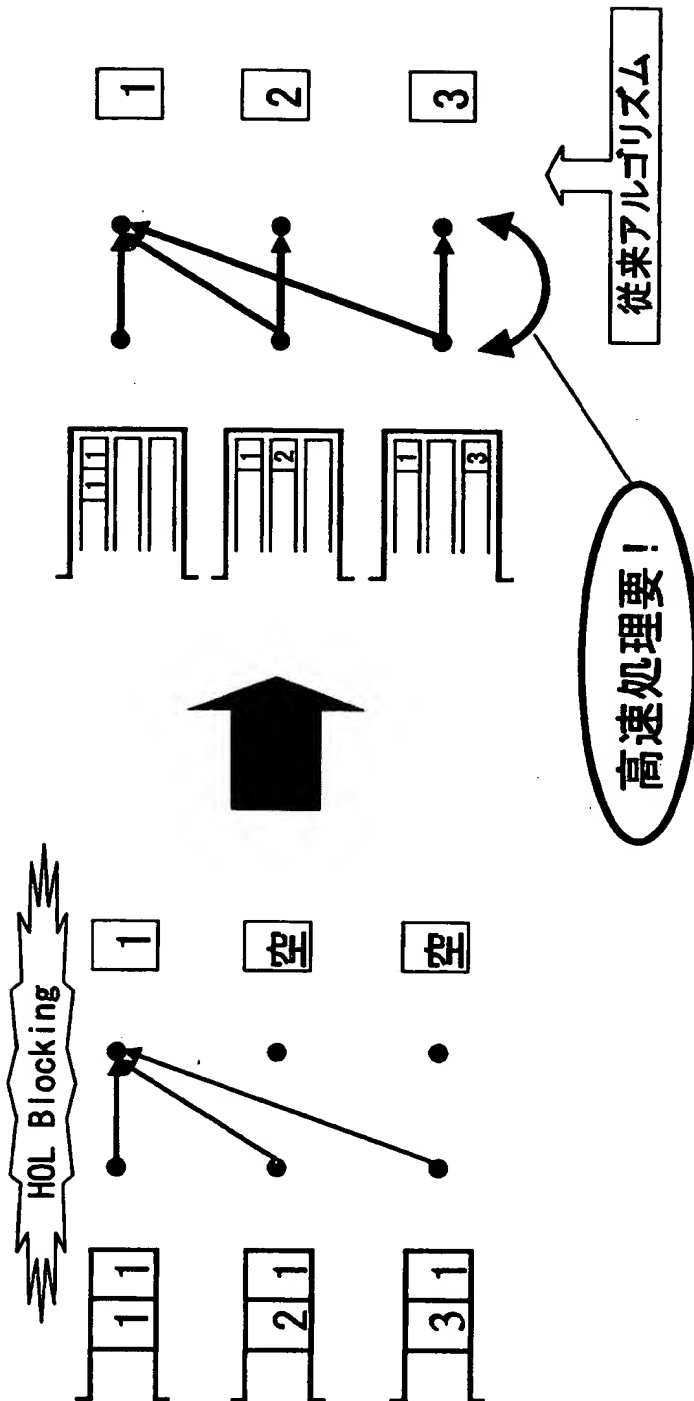
【書類名】 図面

【図 1】



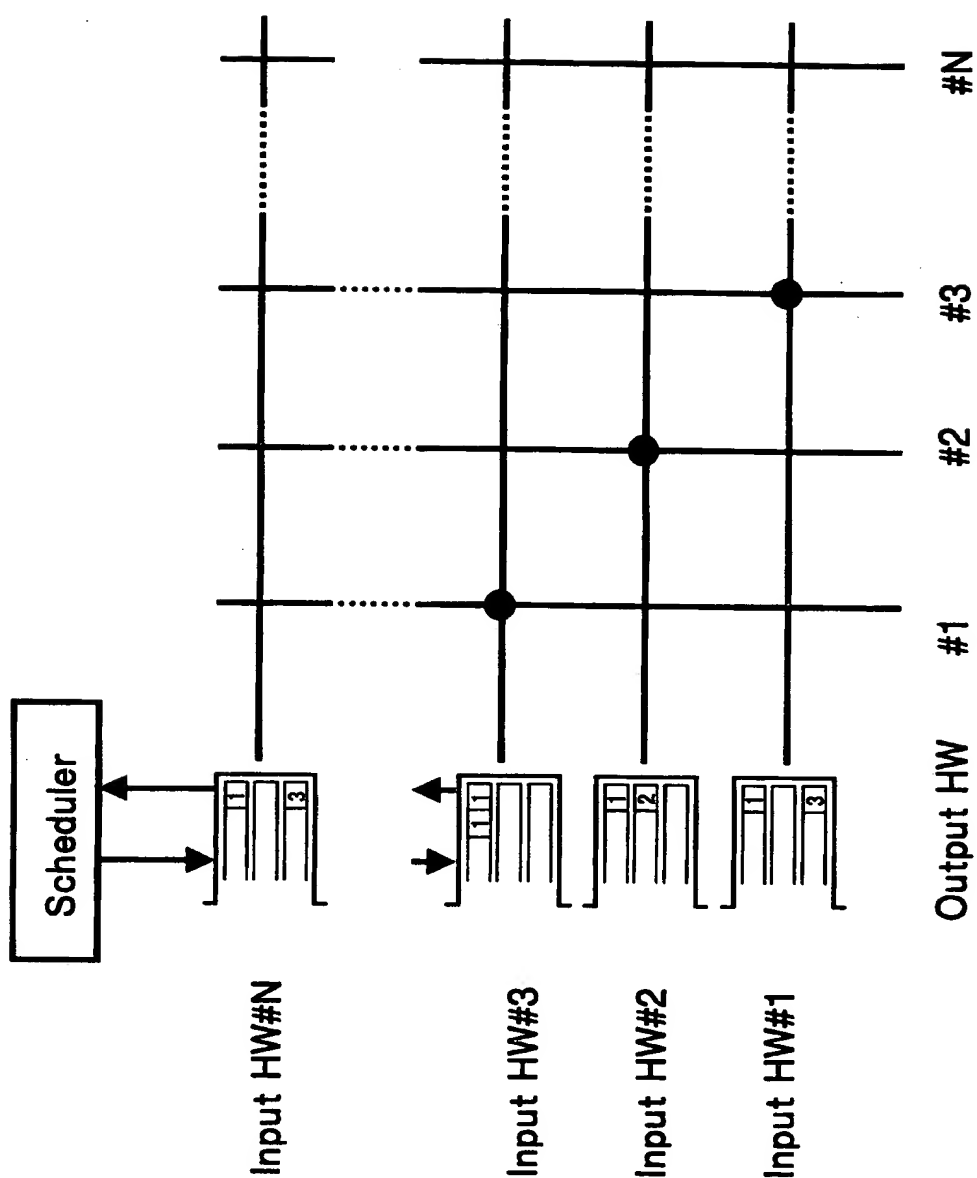
本発明の基本原理図

【図 2】



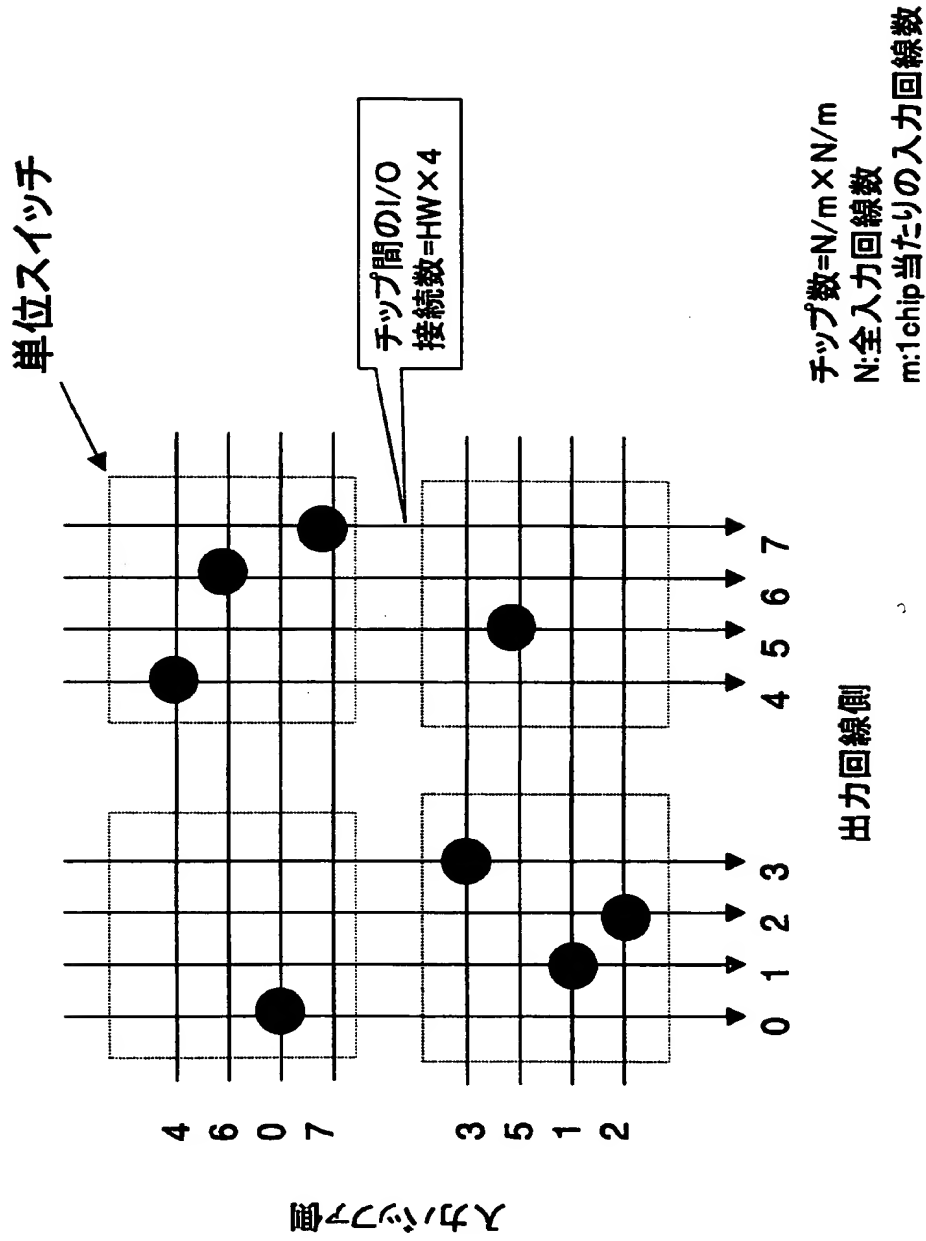
従来技術における入力バッファ型スイッチの問題点を説明するための図

【図 3】



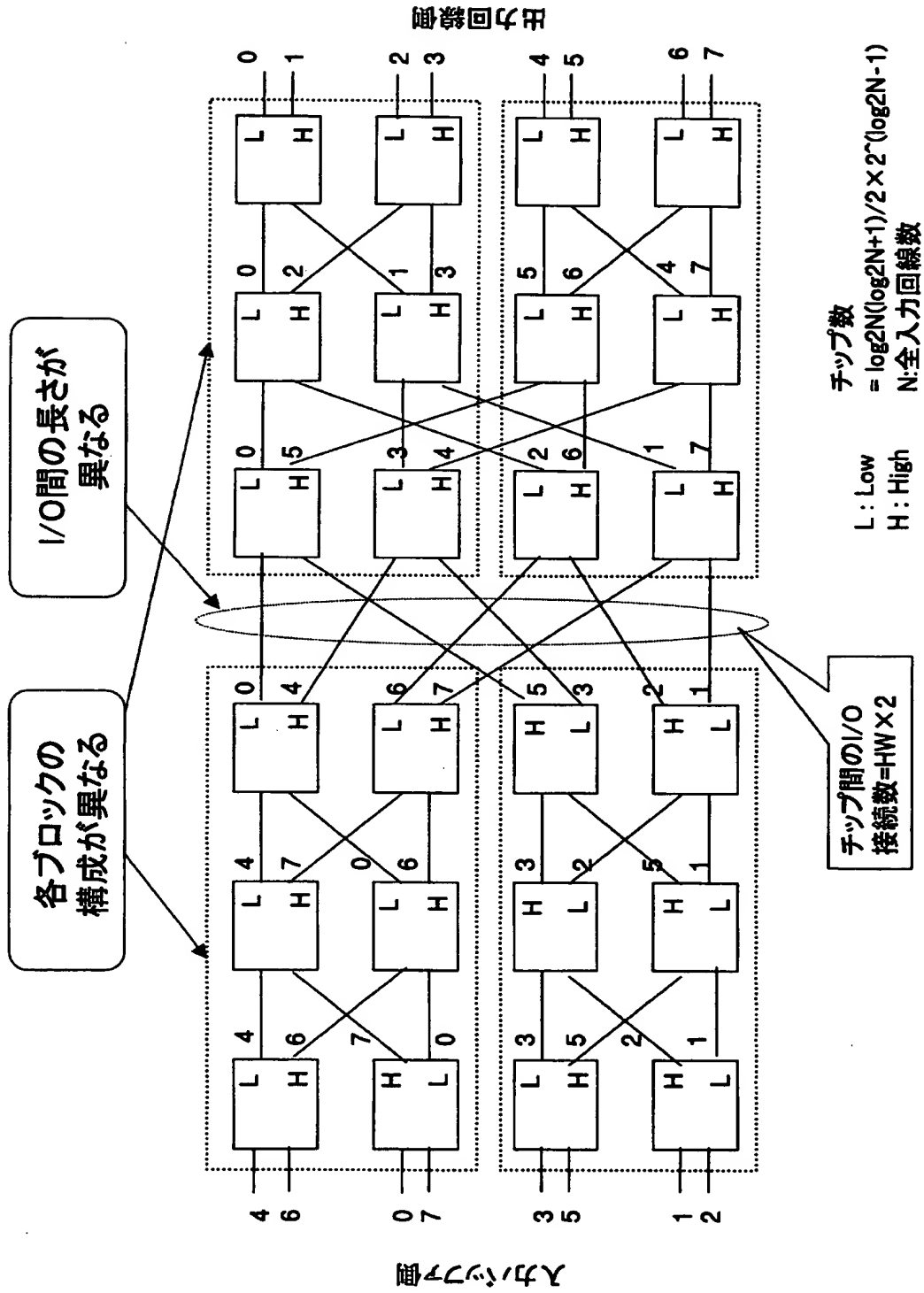
従来技術における入力バッファ型スイッチの構成図

【図 4】



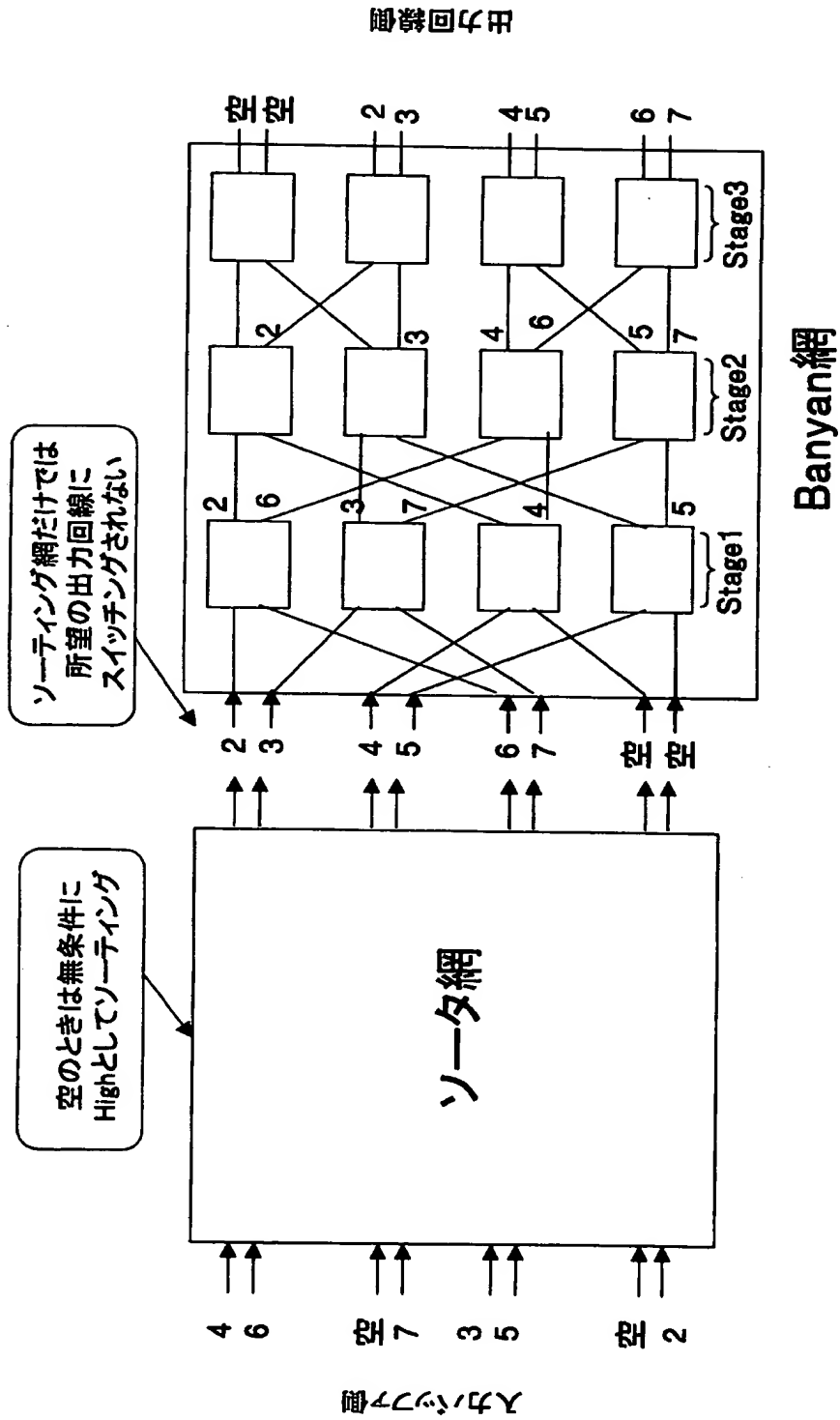
従来技術におけるクロスバー方式の説明

【図 5】



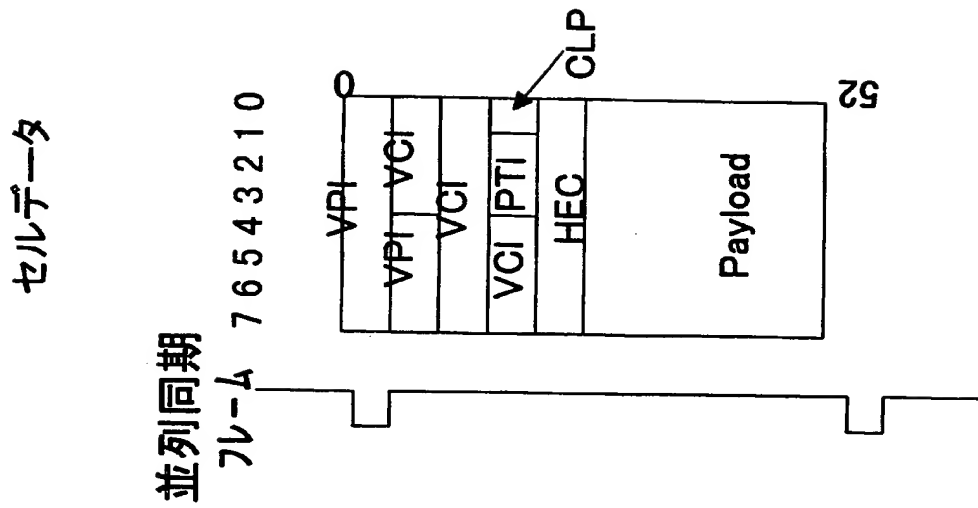
従来技術におけるバッチャー方式の説明図

【図 6】



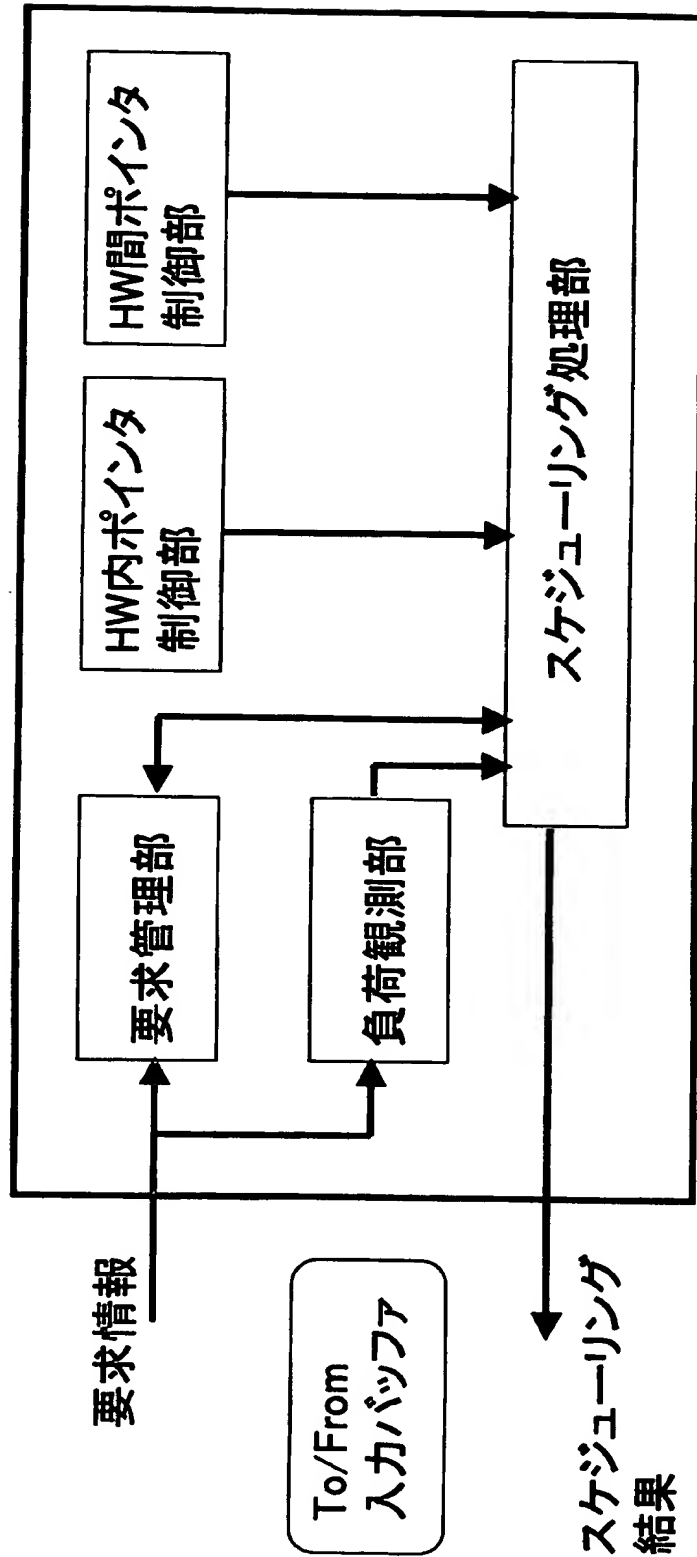
従来技術におけるバンヤン網の説明図

【図 7】



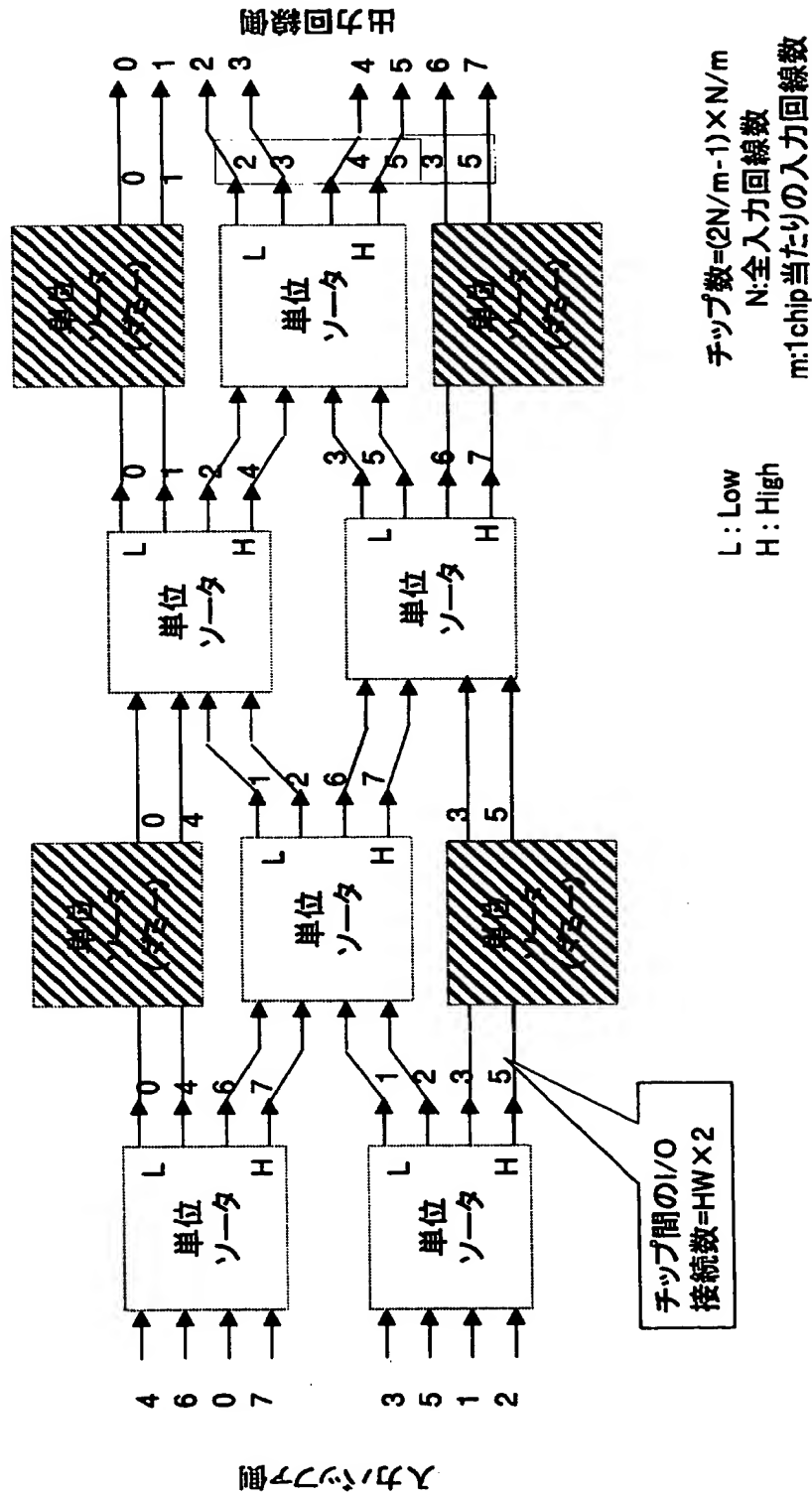
従来技術におけるセルとフレームパルスとを同期させる技術の説明図

【図 8】



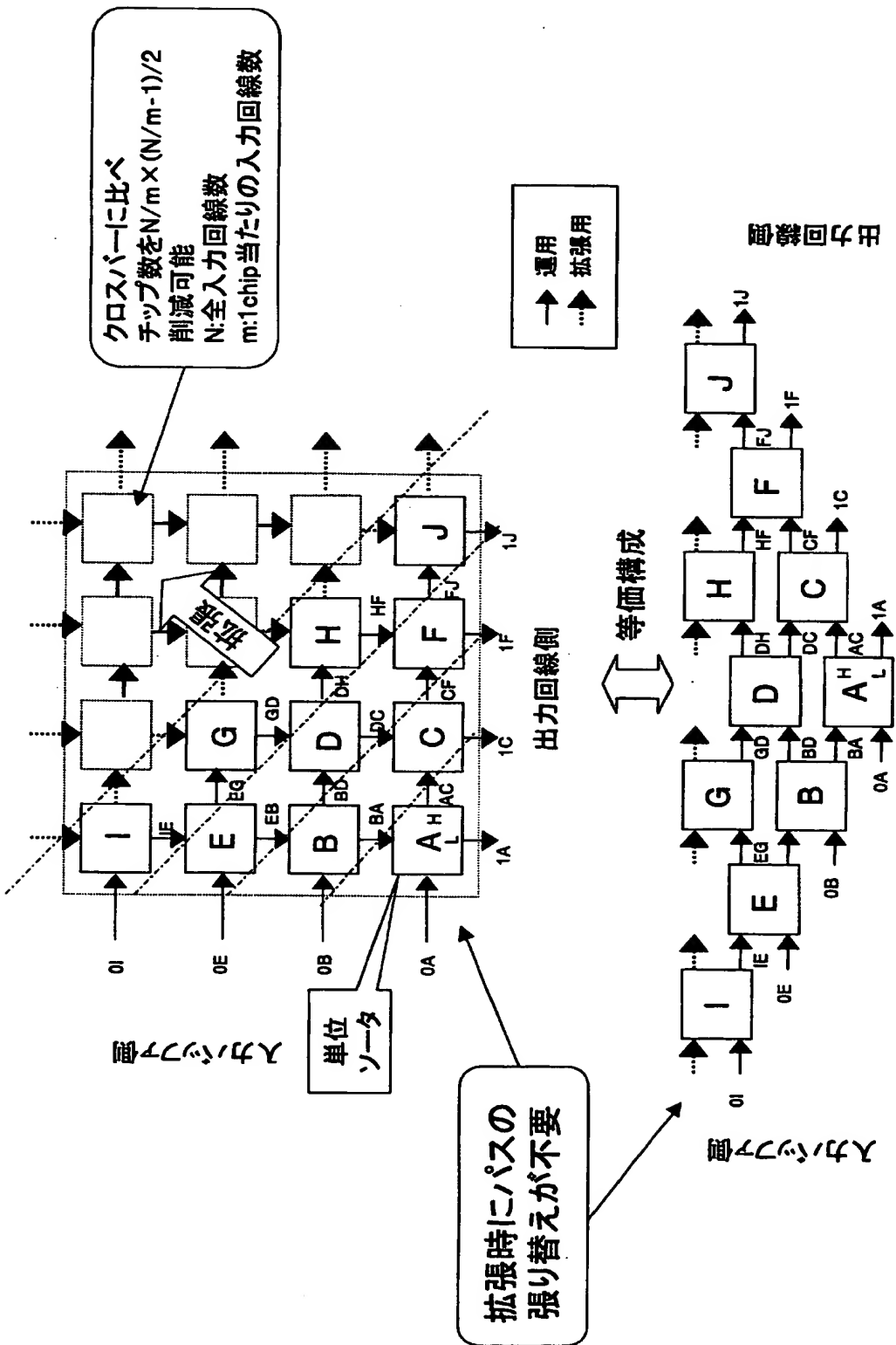
本発明のスケジューラ装置のブロック構成図

【図9】



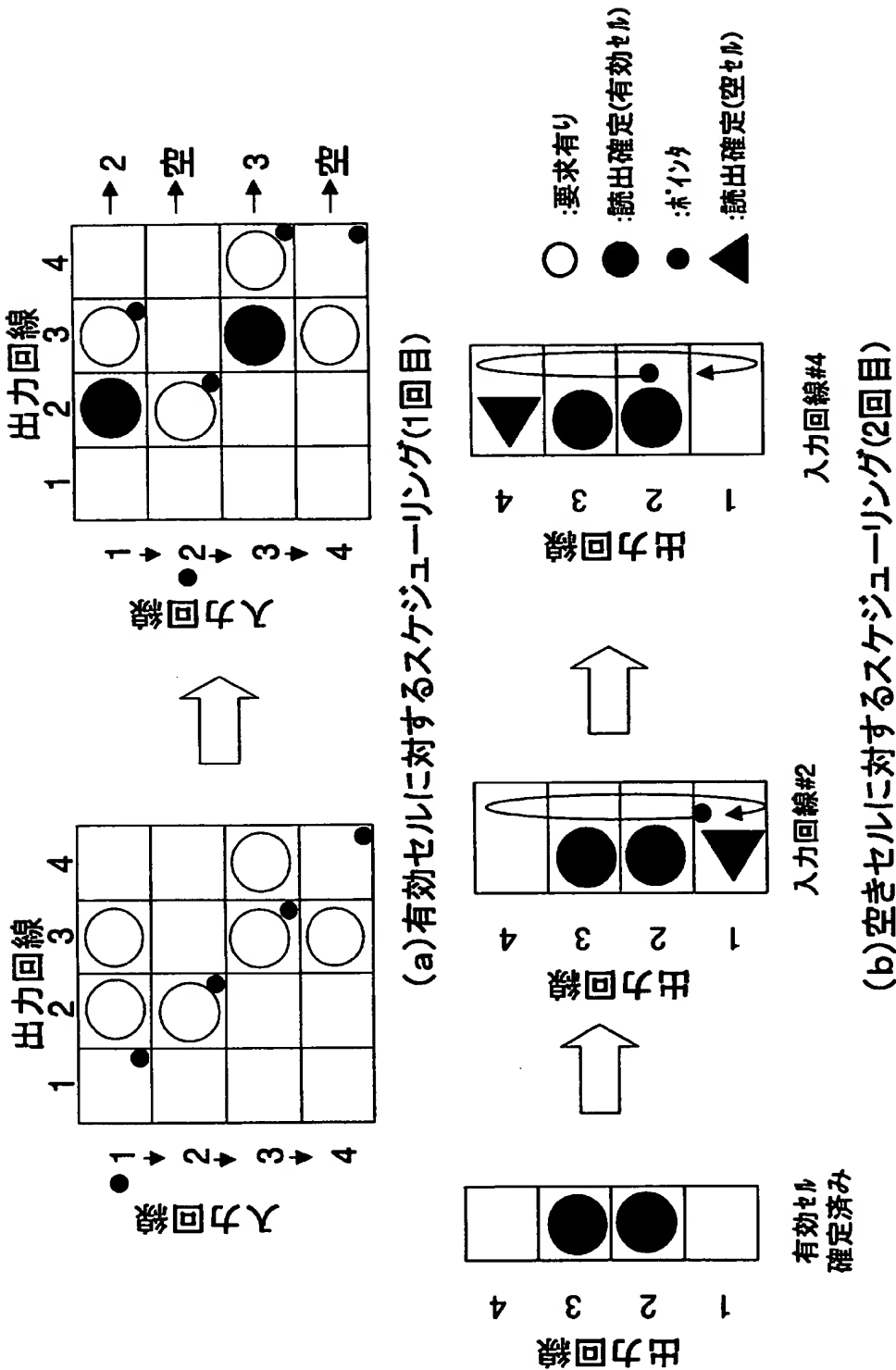
本発明 (請求項 17 に対応) の原理図

【図 10】



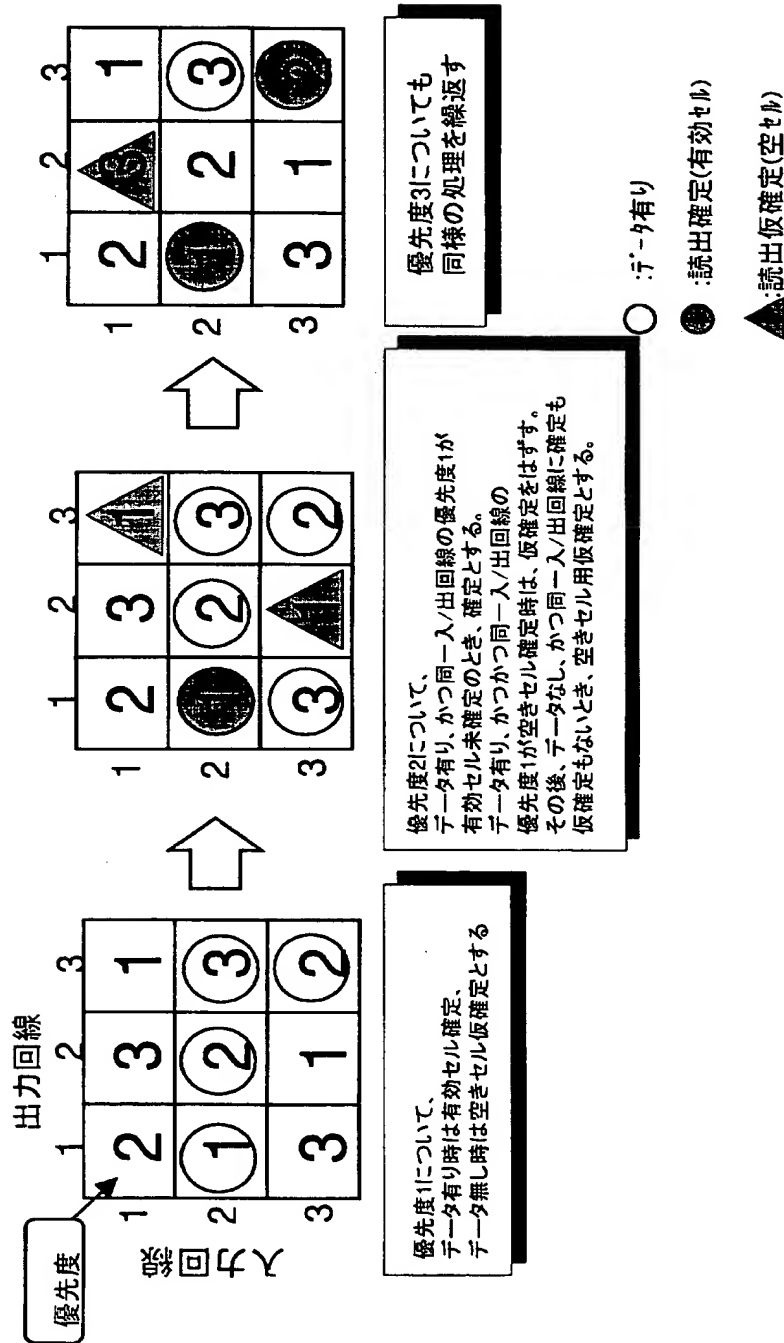
本発明 (請求項 18 に対応) の原理図

【図 1 1】



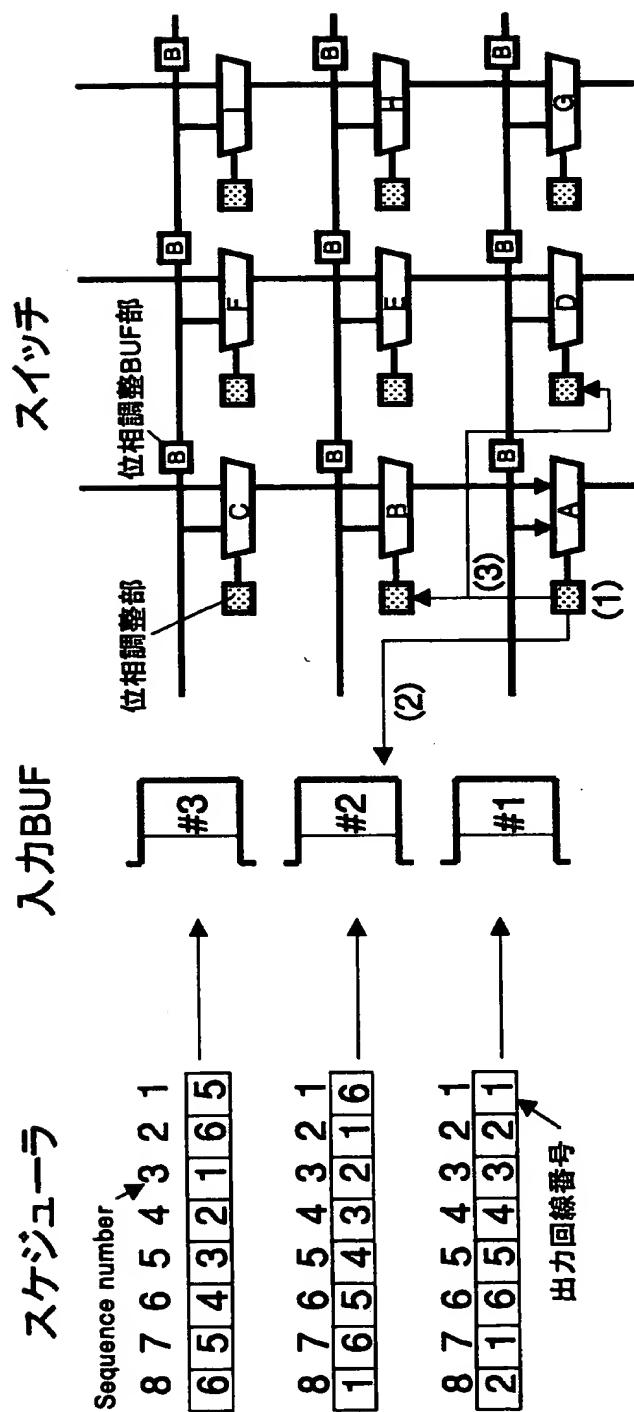
本発明 (請求項 1 9 に対応) の原理図

【図 12】



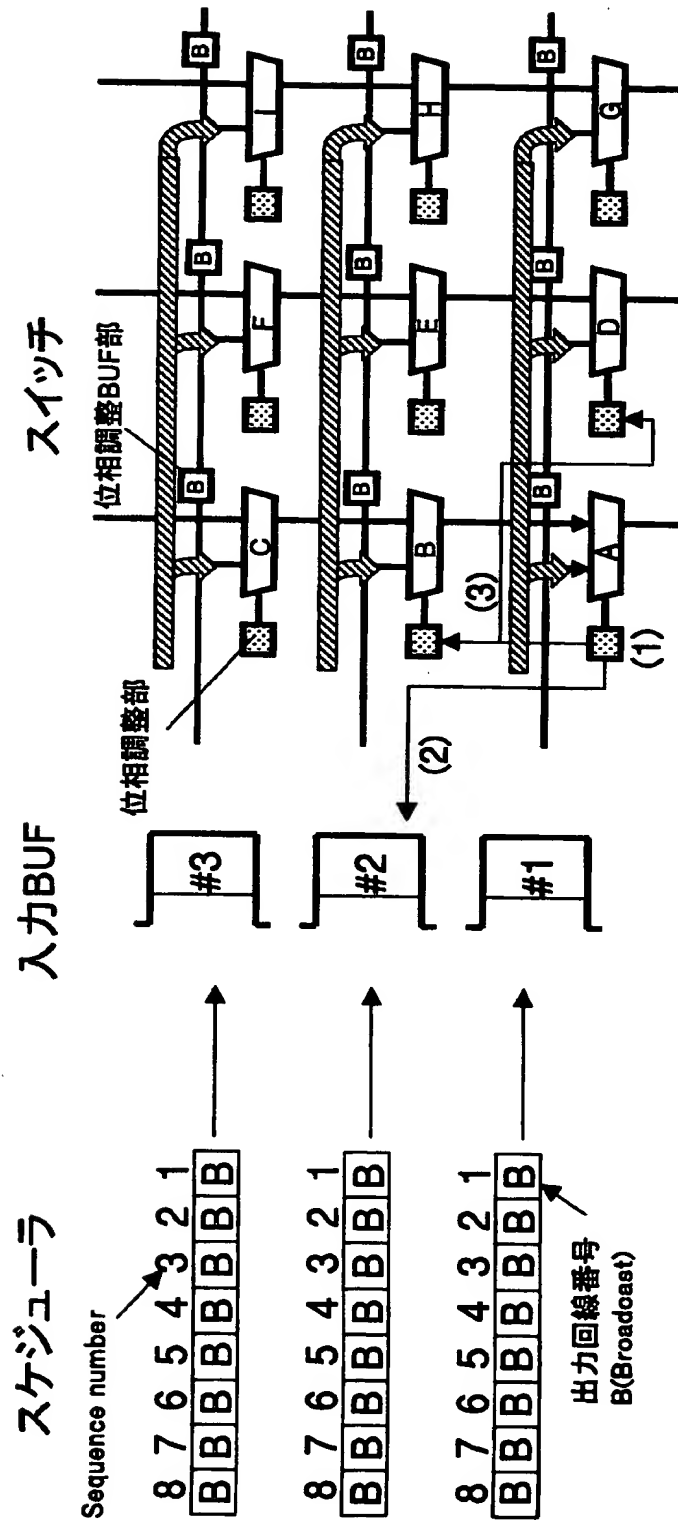
本発明 (請求項 20 に対応) の原理図

【図 13】



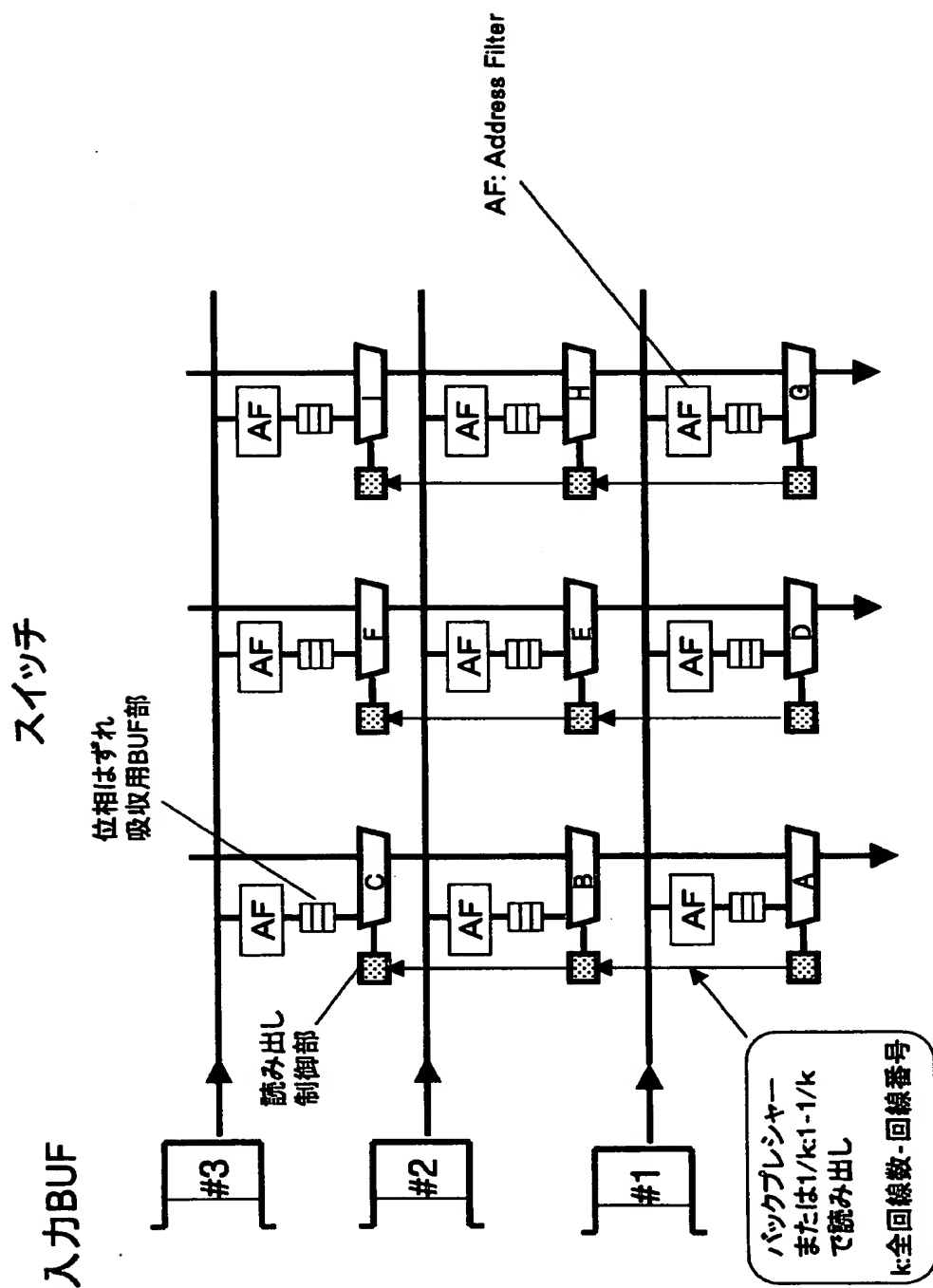
本発明 (請求項 21 に対応) の原理図

【図 15】



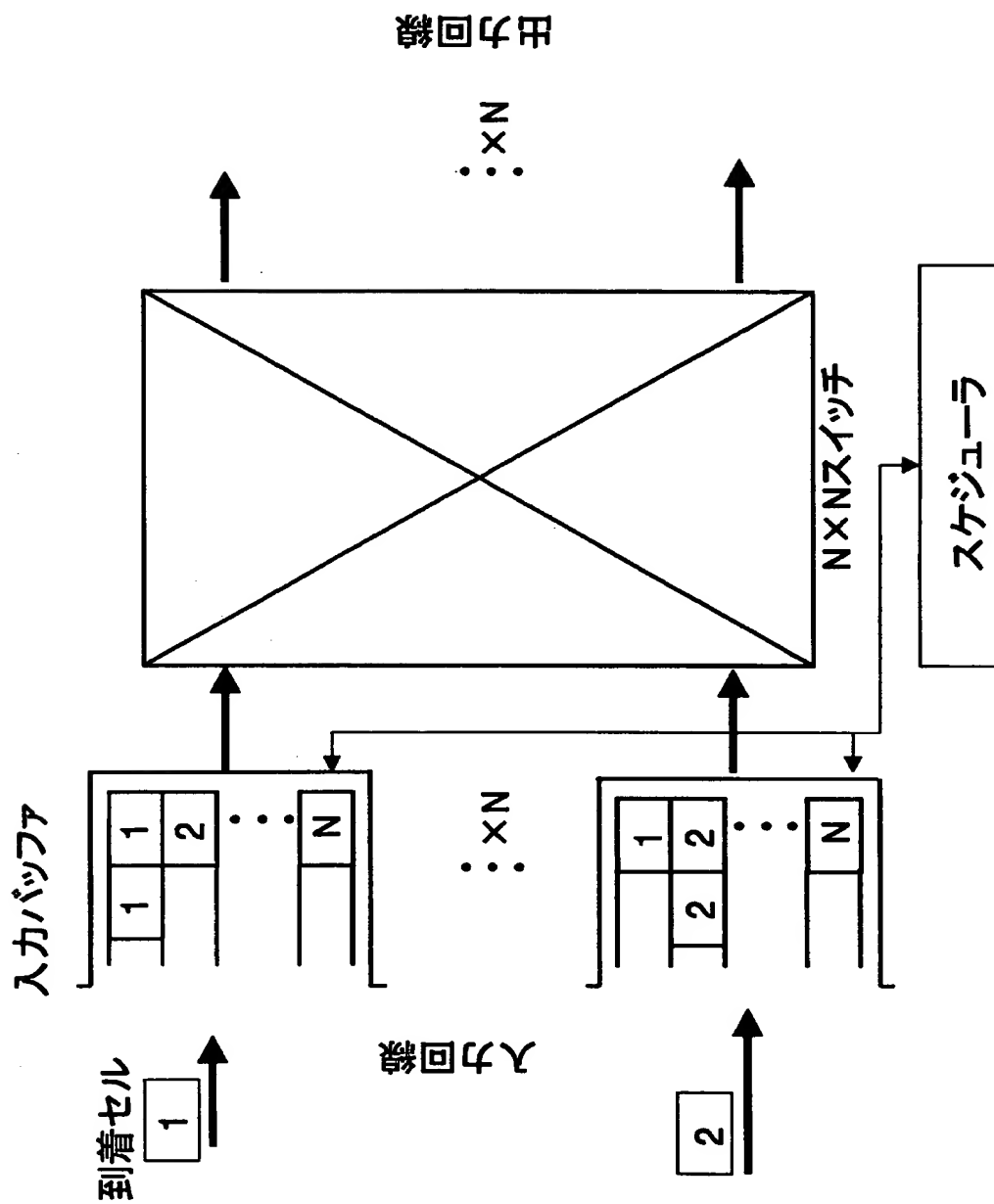
本発明 (請求項 22 に対応) の原理図

【图 16】



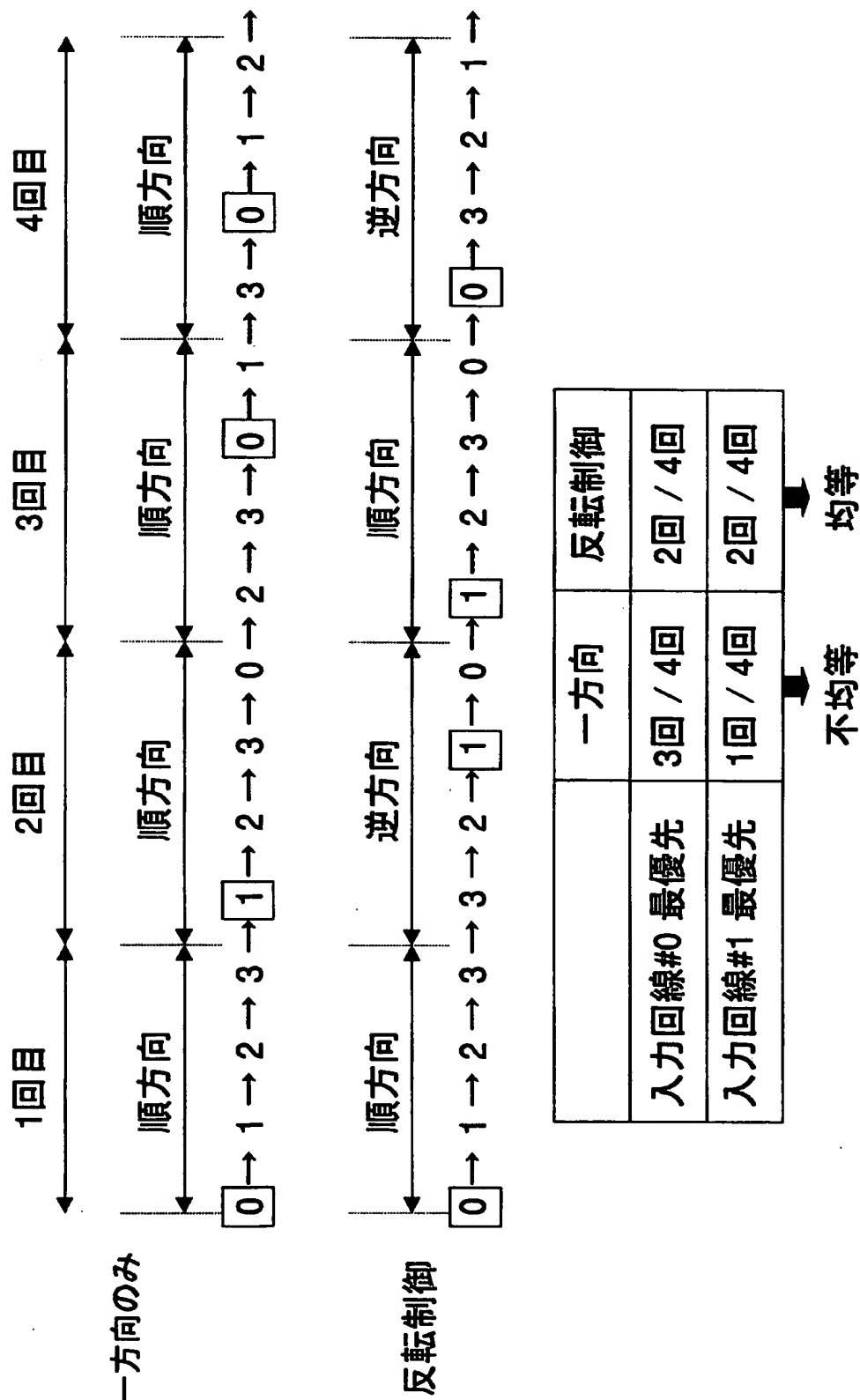
本発明（請求項 23 に対応）の原理図

【図 17】



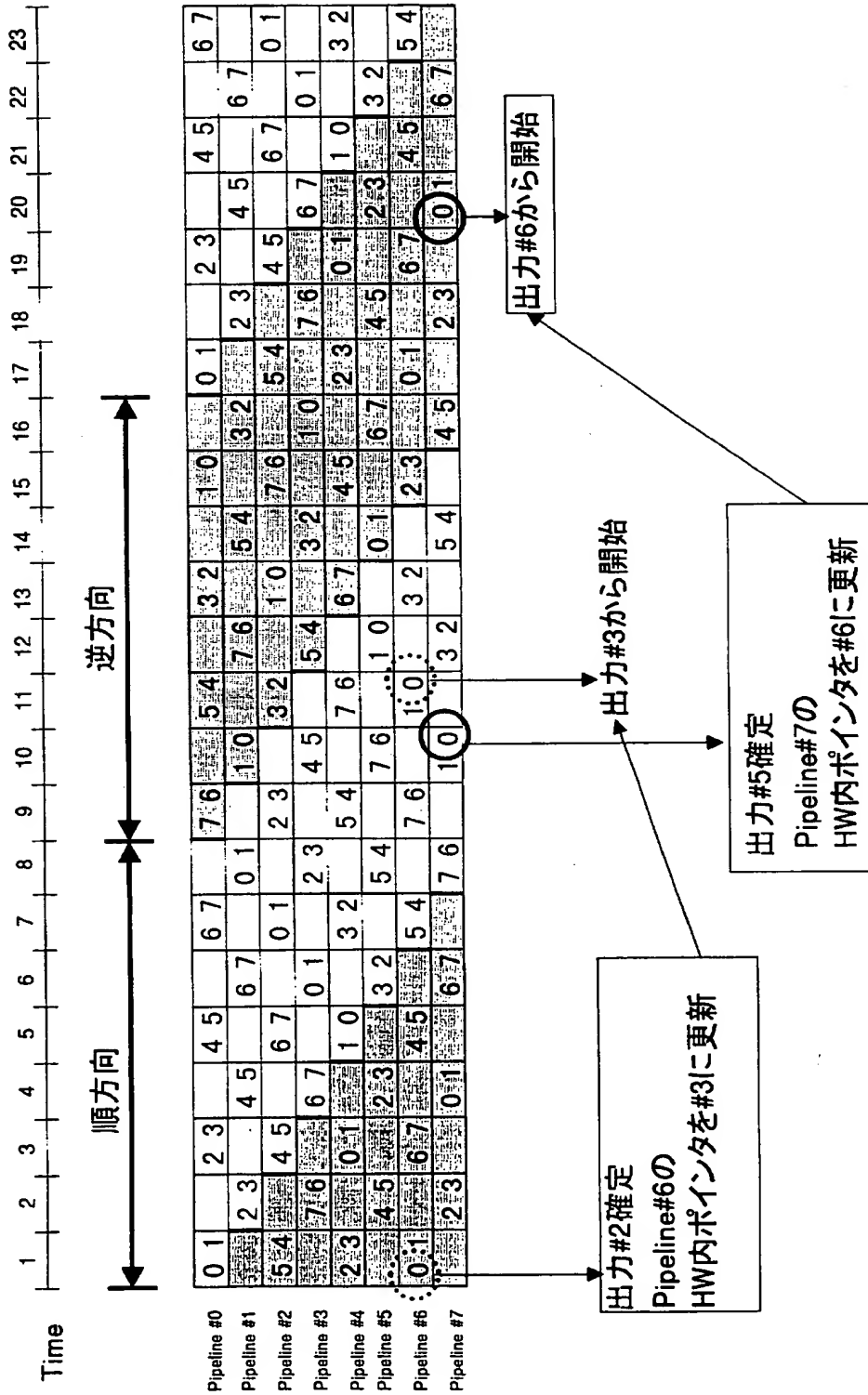
実施例 1 のシステム構成を示すブロック図

【図 18】



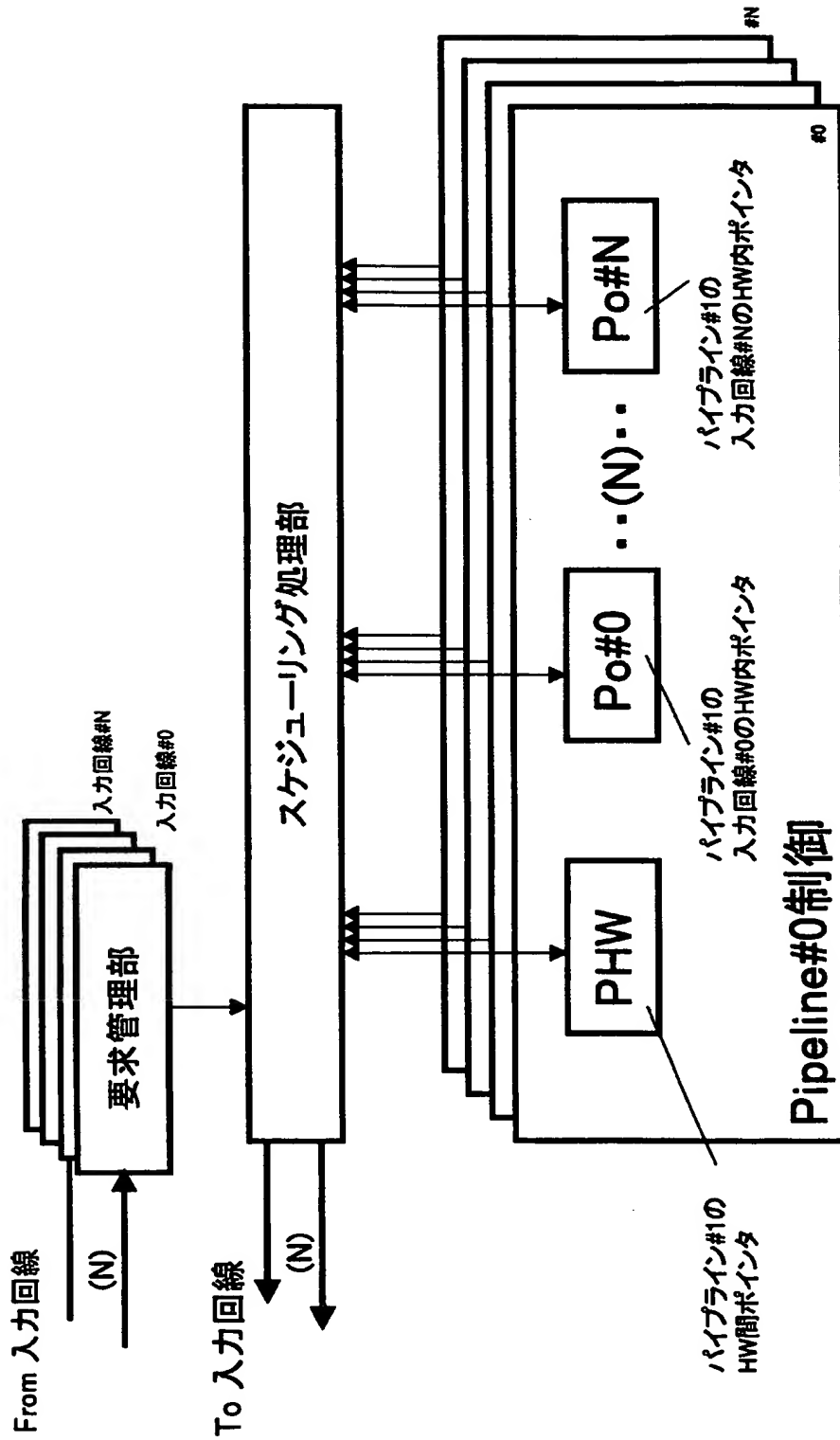
実施例 1 におけるハイウェイ間ポイントの反転制御を説明するための図

【図 19】



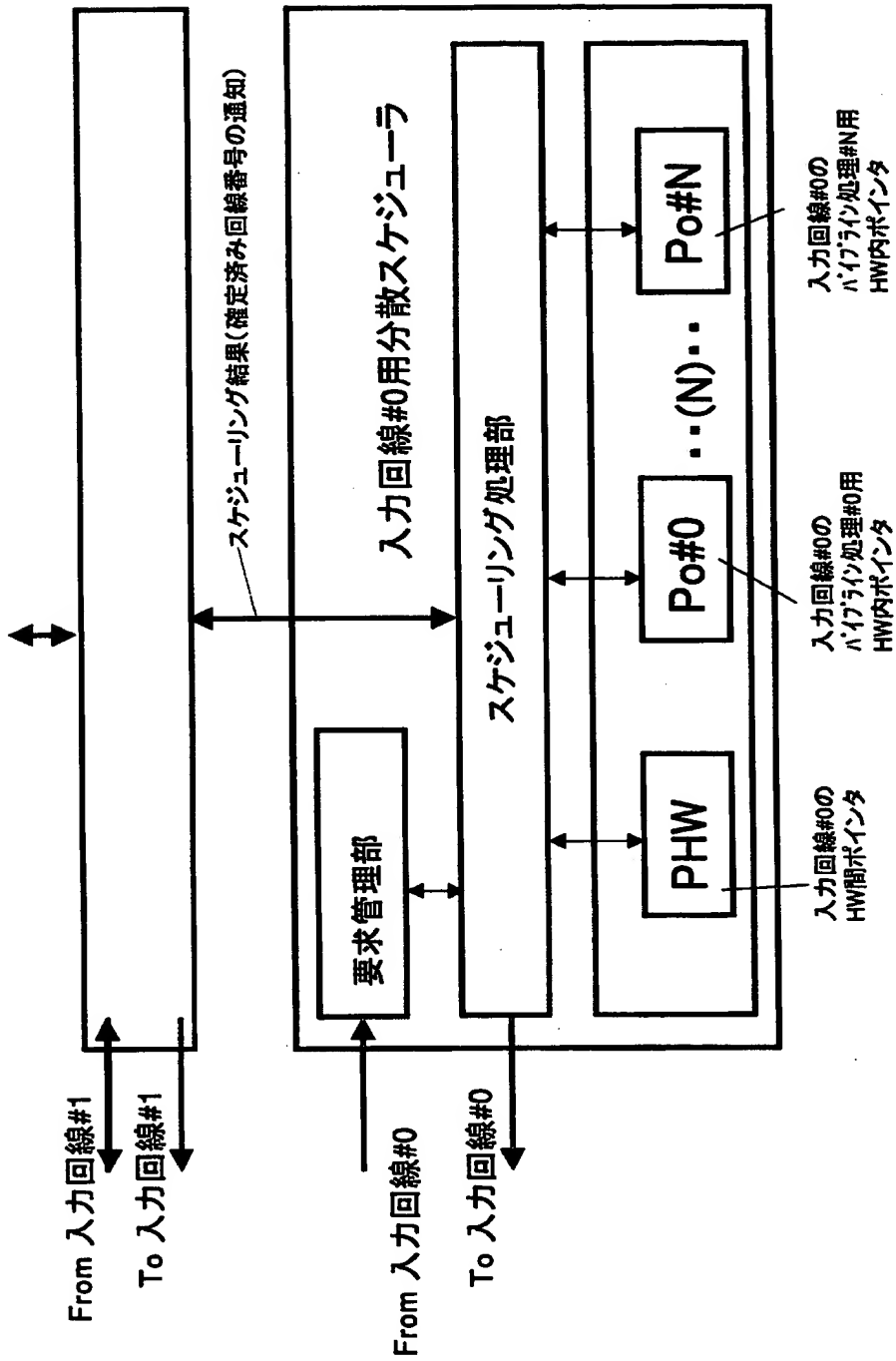
実施例 2 におけるパイプライン処理時のポインタ制御を説明するための図

【図 20】



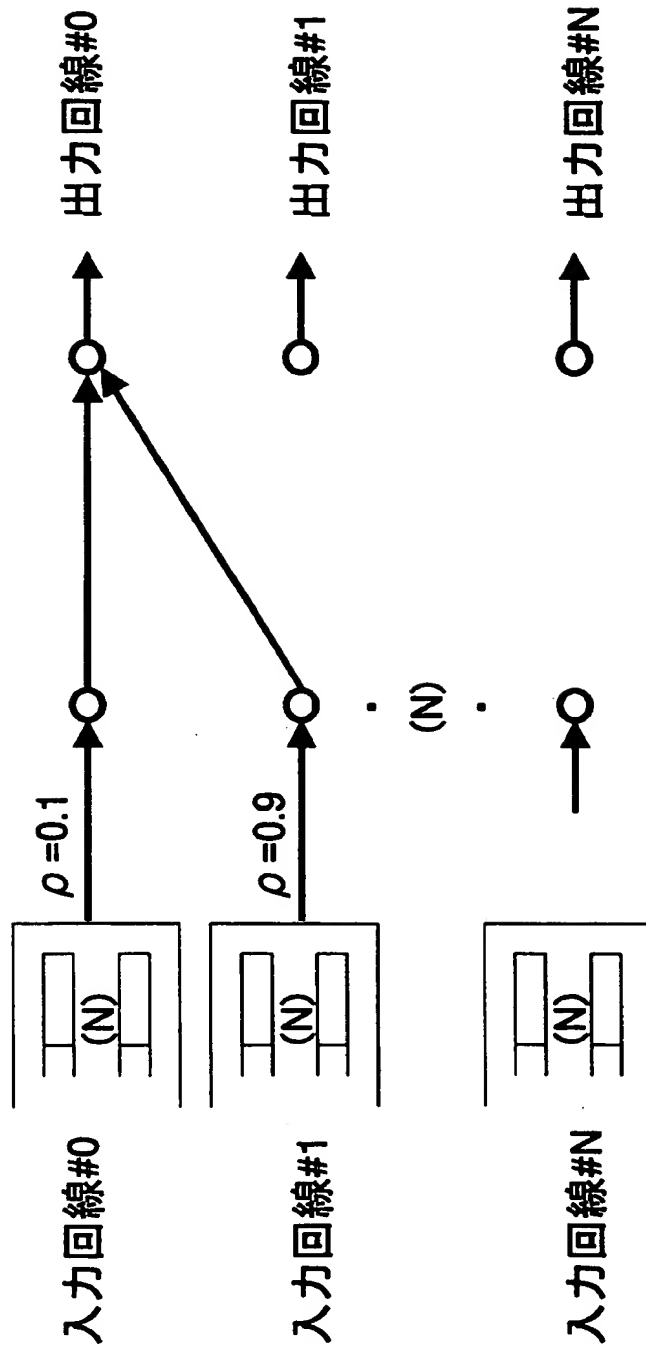
実施例 2 におけるスケジューリング処理部を集中背部した場合のブロック図

【図 2 1】



実施例 2 におけるスケジューリング処理部を分散配備した場合のブロック図

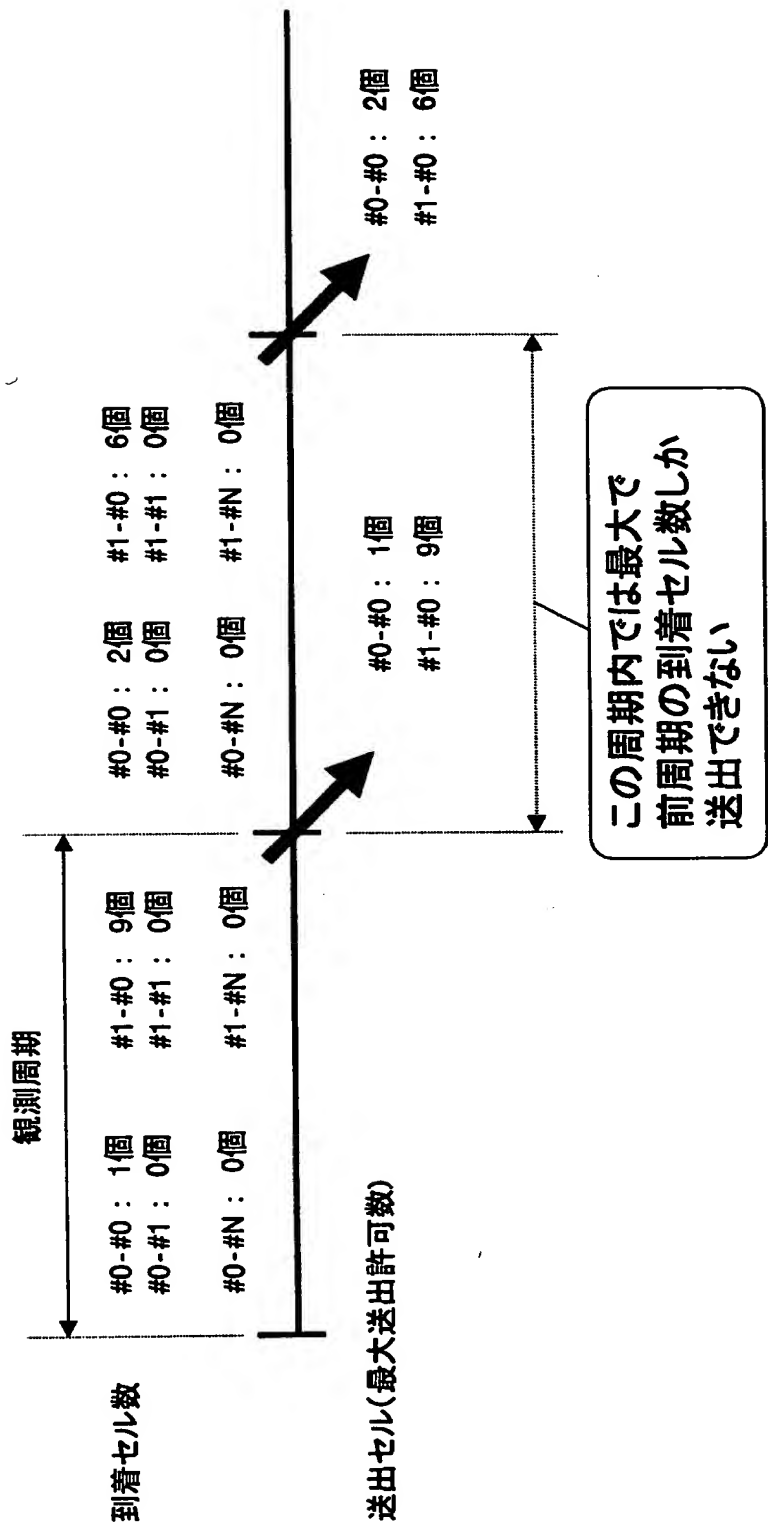
【図 22】



パス #0-#0の選択確率:0.5 → 入力負荷0.1 → 問題なし
 パス #1-#0の選択確率:0.5 → 入力負荷0.9 → キュー長増加

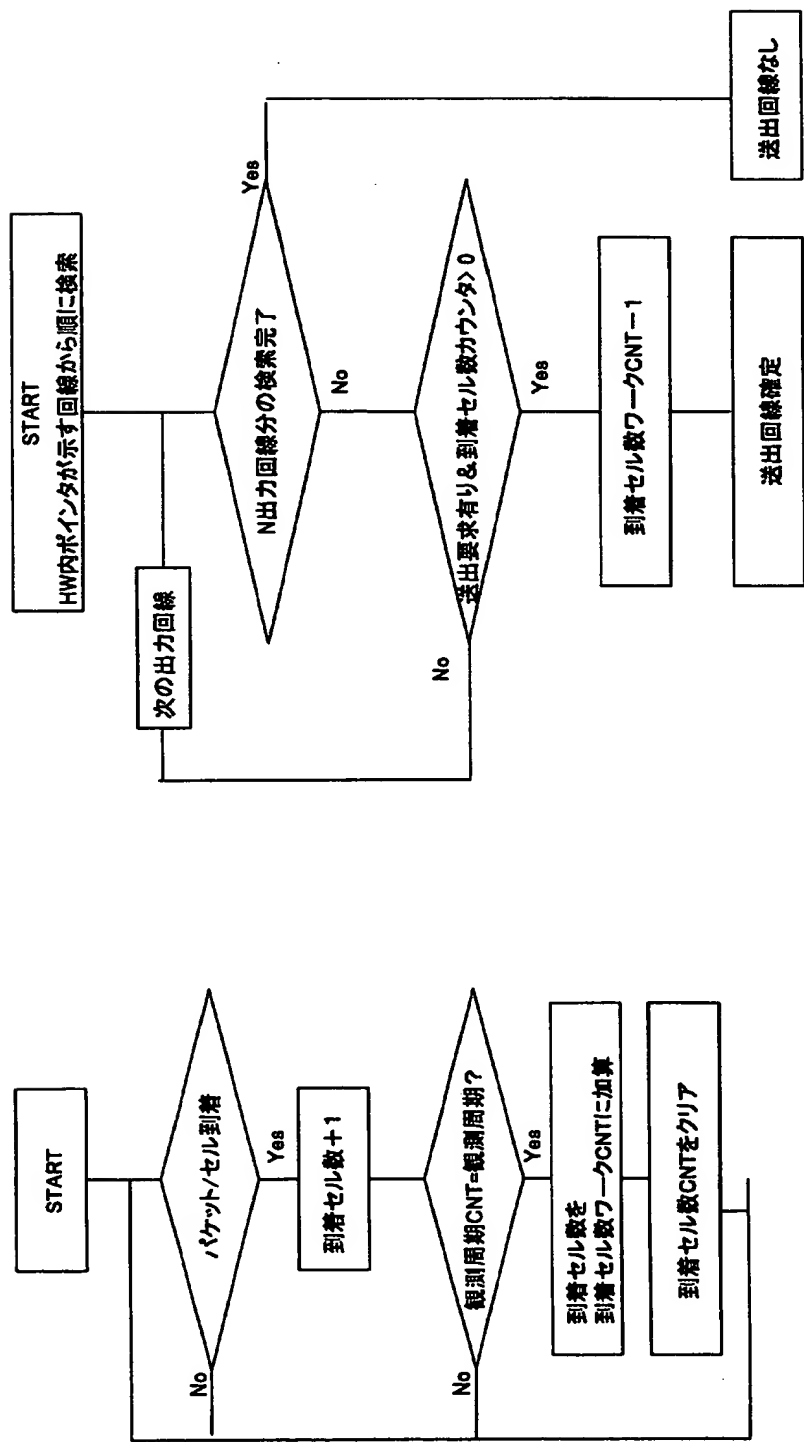
実施例 3 における不均等負荷時における入力回線と出力回線との関係を示す説明図

【図 23】

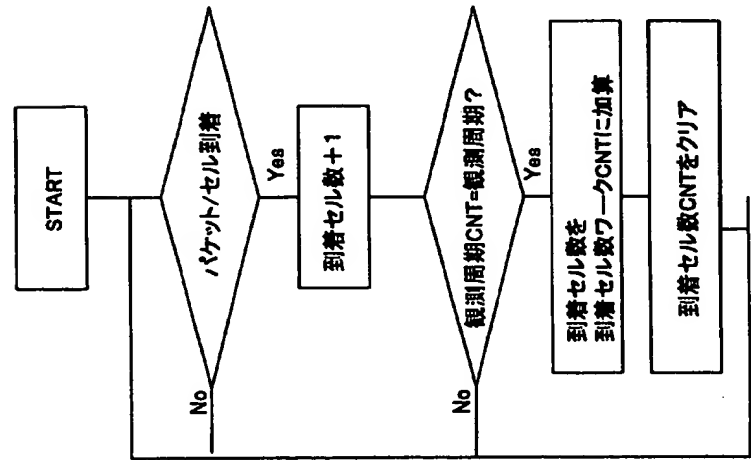


実施例 3 における送出セル数を到着セル数に制限する場合のフロー図

【 図 2 4 】



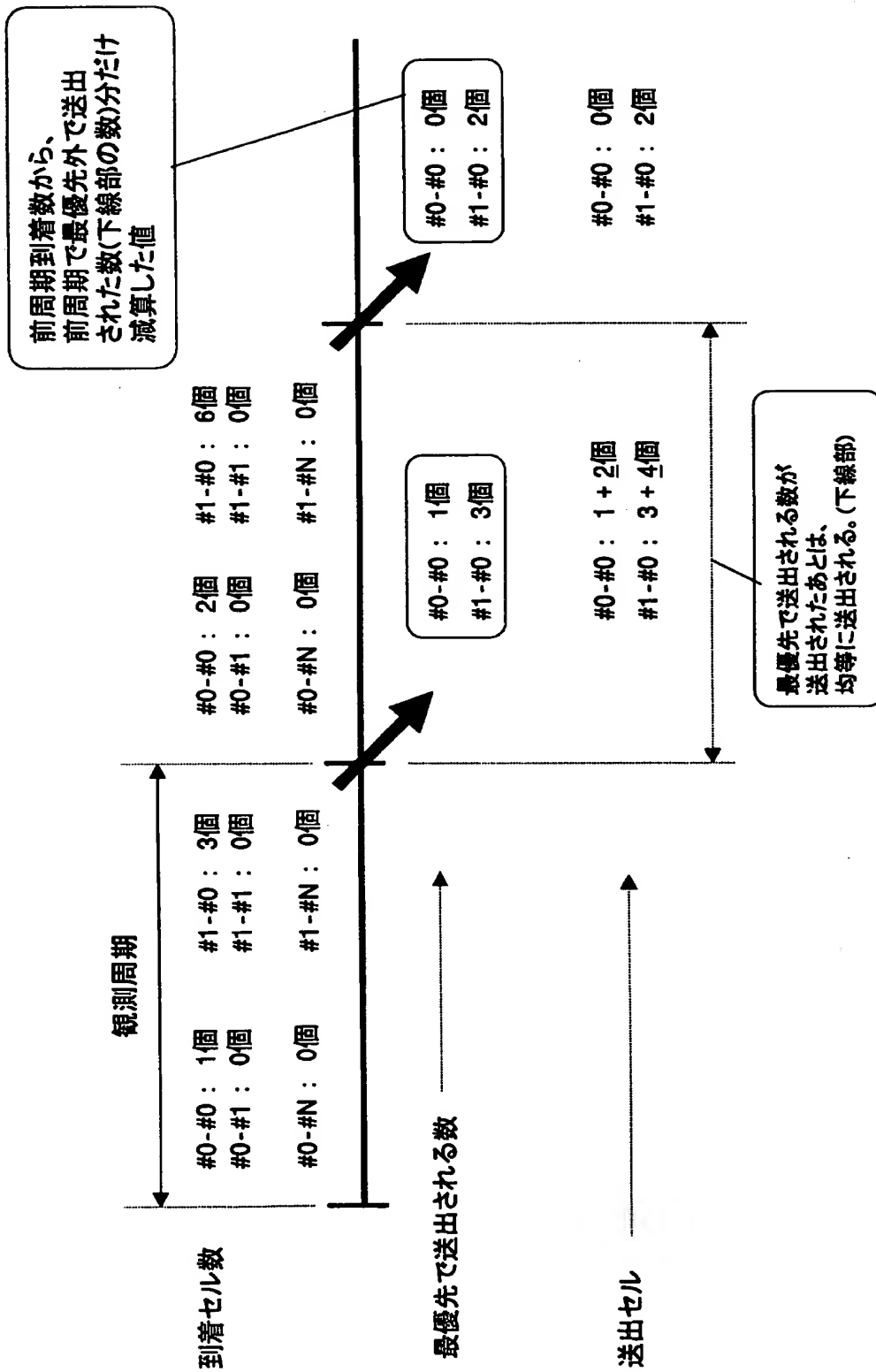
(b) スケジューリングフロー



(a) 負荷観測フロー

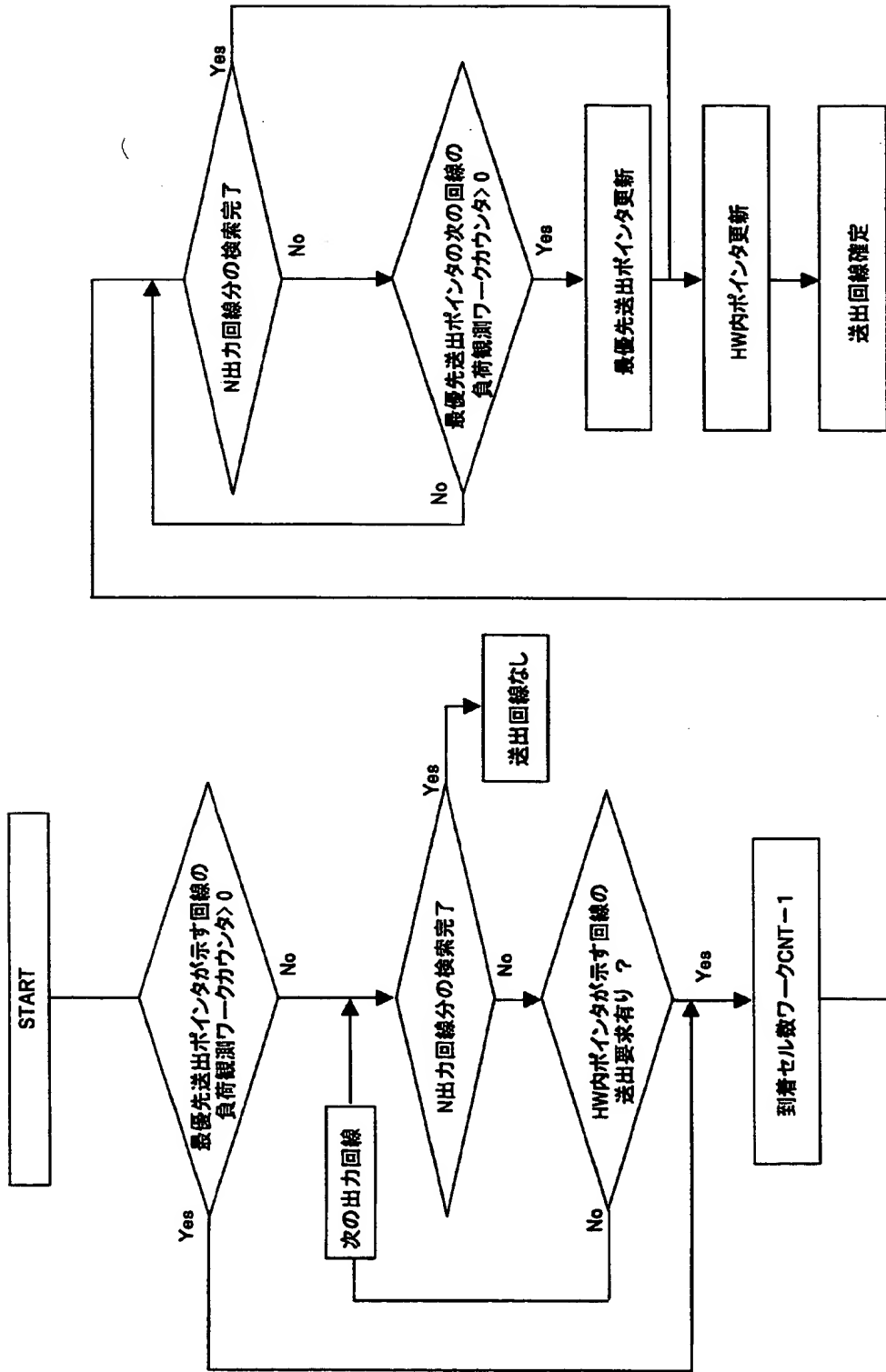
実施例 3 における送出セル数を規制する場合のフロー図

【図 25】



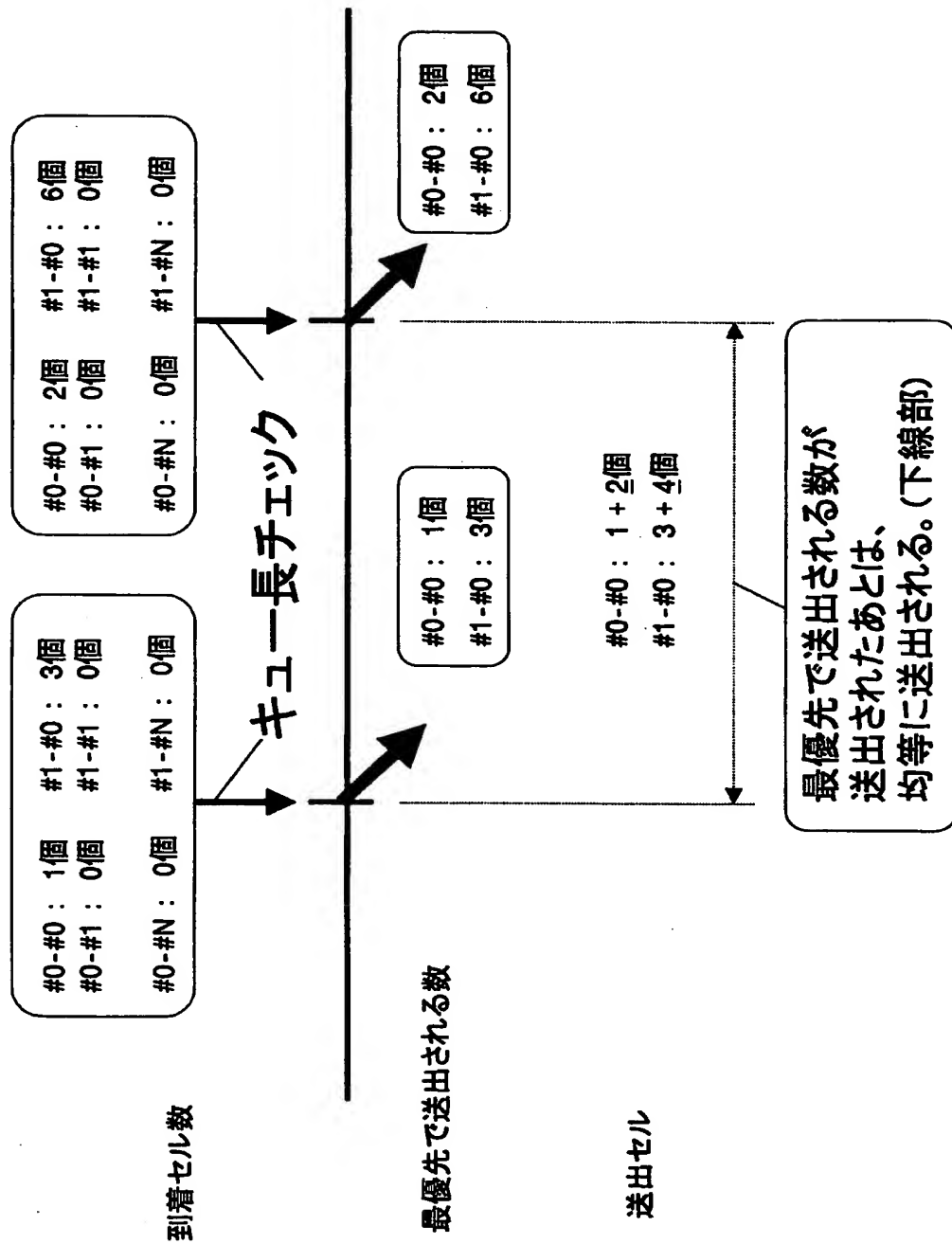
実施例 4 における最優先送出回線を到着セル数の割合で決定する方法を説明する図

【図 2 6】



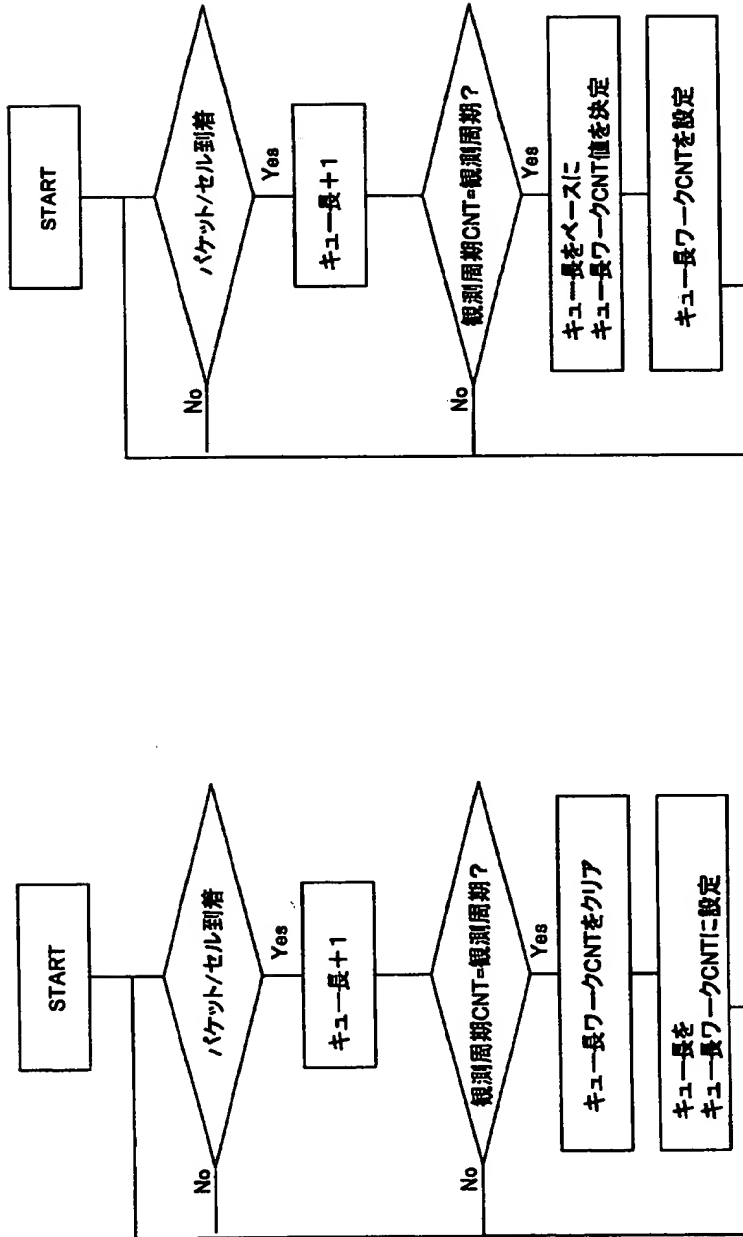
実施例 4 における最優先送出ポイントを用いたスケジューリングフロー図

【図 27】



実施例 5 においてキュー長を用いた最優先送出回線決定方法を説明するための図

【図 2 8】



(a) キュー長ベースの負荷観測フロー(その1)

(b)(c) キュー長ベースの負荷観測フロー(その2)

実施例 5 におけるキュー長ベースの負荷観測のためのフロー図

【図 29】

a) 測定結果をそのまま設定

パス#0-パス#0キュー長: 10個 → 10
 パス#1-パス#0キュー長: 2個 → 1
 パス#2-パス#0キュー長: 5個 → 3
 パス#3-パス#0キュー長: 10個 → 10

b) 各キュー長で次周期の割合で送出割合決定(例: 一定周期=20パケット時間)

パス#0-パス#0キュー長: 10個 → $10 / (10+2+5+10) * 20 = 7.4 = 7$
 パス#1-パス#0キュー長: 2個 → $2 / (10+2+5+10) * 20 = 1.4 = 1$
 パス#2-パス#0キュー長: 5個 → $5 / (10+2+5+10) * 20 = 3.7 = 3$
 パス#3-パス#0キュー長: 10個 → $10 / (10+2+5+10) * 20 = 7.4 = 7$

c) キュー長に応じて次周期の送出割合決定(例: 一定周期=20パケット時間)

キュー長: 1~3 → 0.05 * 20 → 1
 キュー長: 3~10 → 0.1 * 20 → 2
 キュー長: 10~20 → 0.2 * 20 → 4
 キュー長: 20~30 → 0.3 * 20 → 6

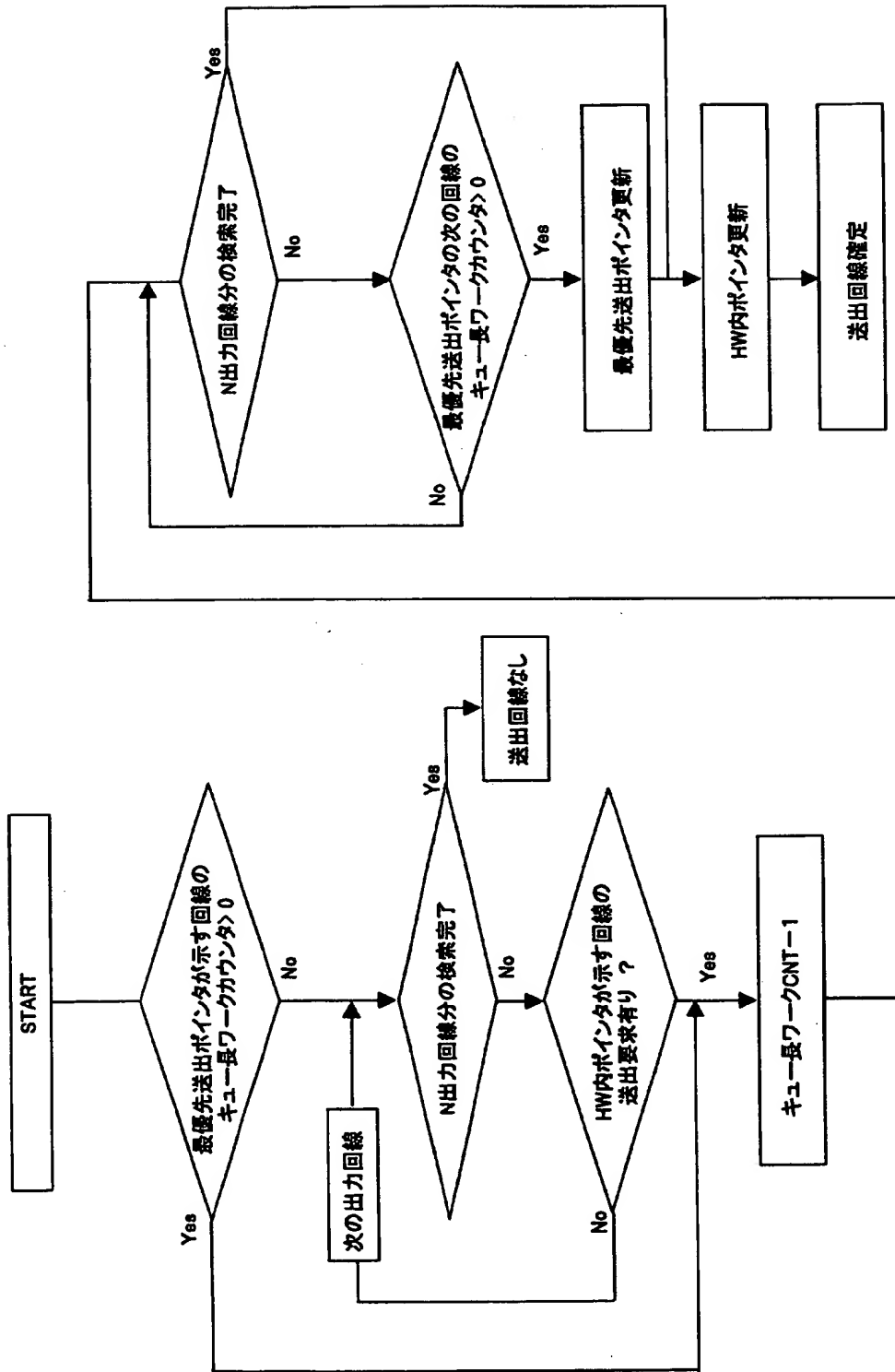
← 範囲および割合は任意に設定可能

キュー長: 90~100 → 1.0 * 20 → 20

パス#0-パス#0キュー長: 10個 → 2
 パス#1-パス#0キュー長: 2個 → 1
 パス#2-パス#0キュー長: 5個 → 2
 パス#3-パス#0キュー長: 10個 → 2

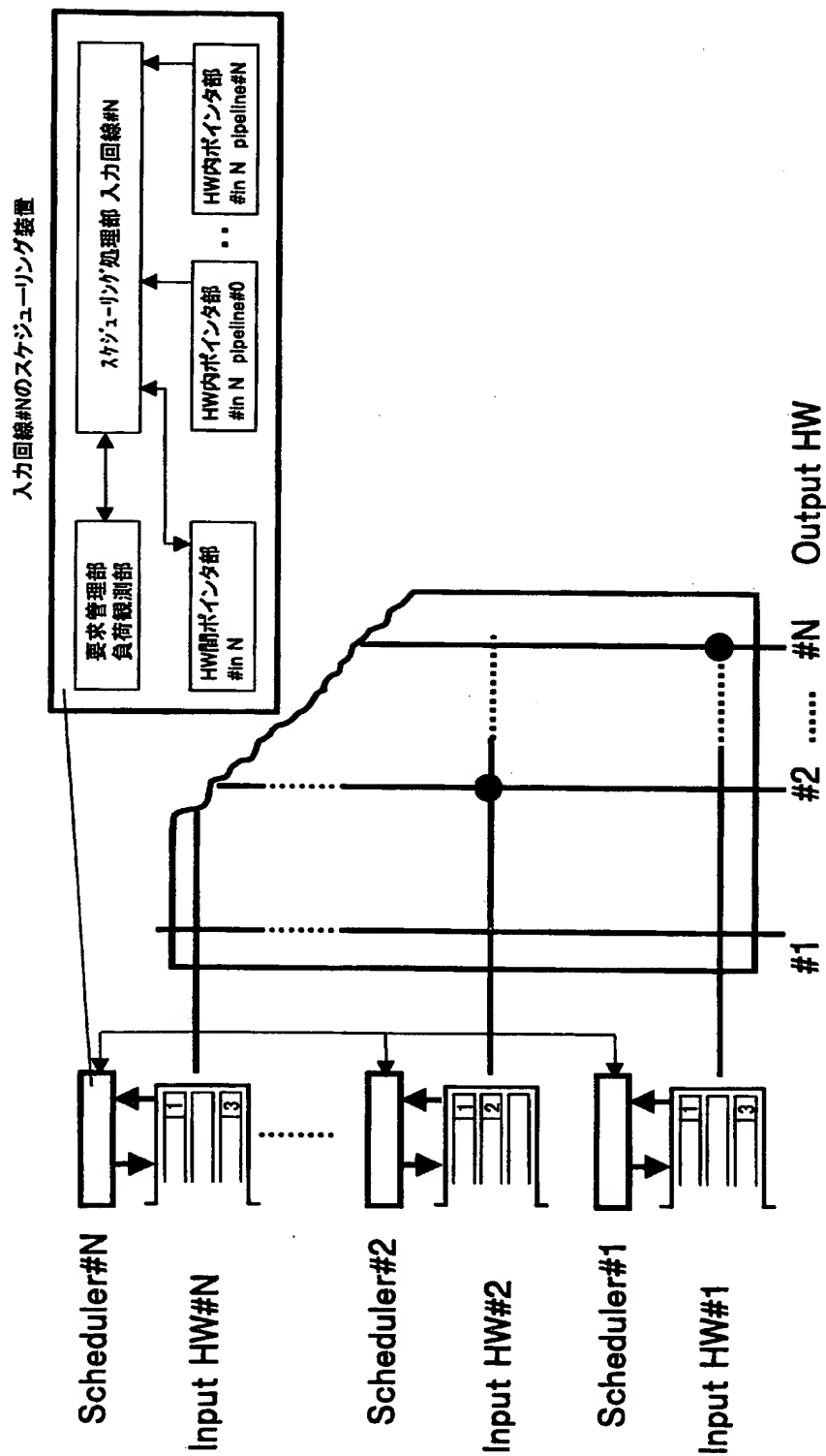
実施例 5 においてキュー長ワークカウンタ設定値の決定方法についての説明図

【図 3 0】



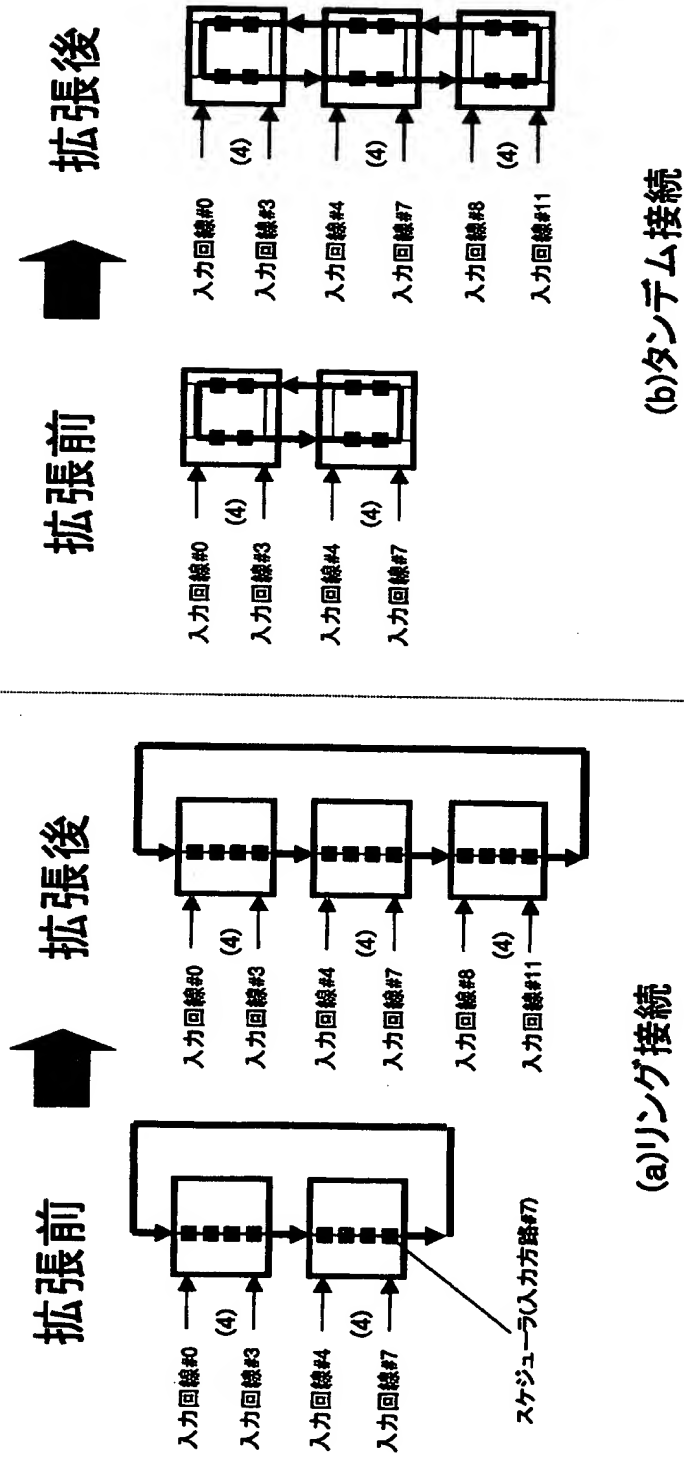
実施例5におけるキュー長ベースの最優先送出ポイントを用いた制御フロー図

【図 3 1】



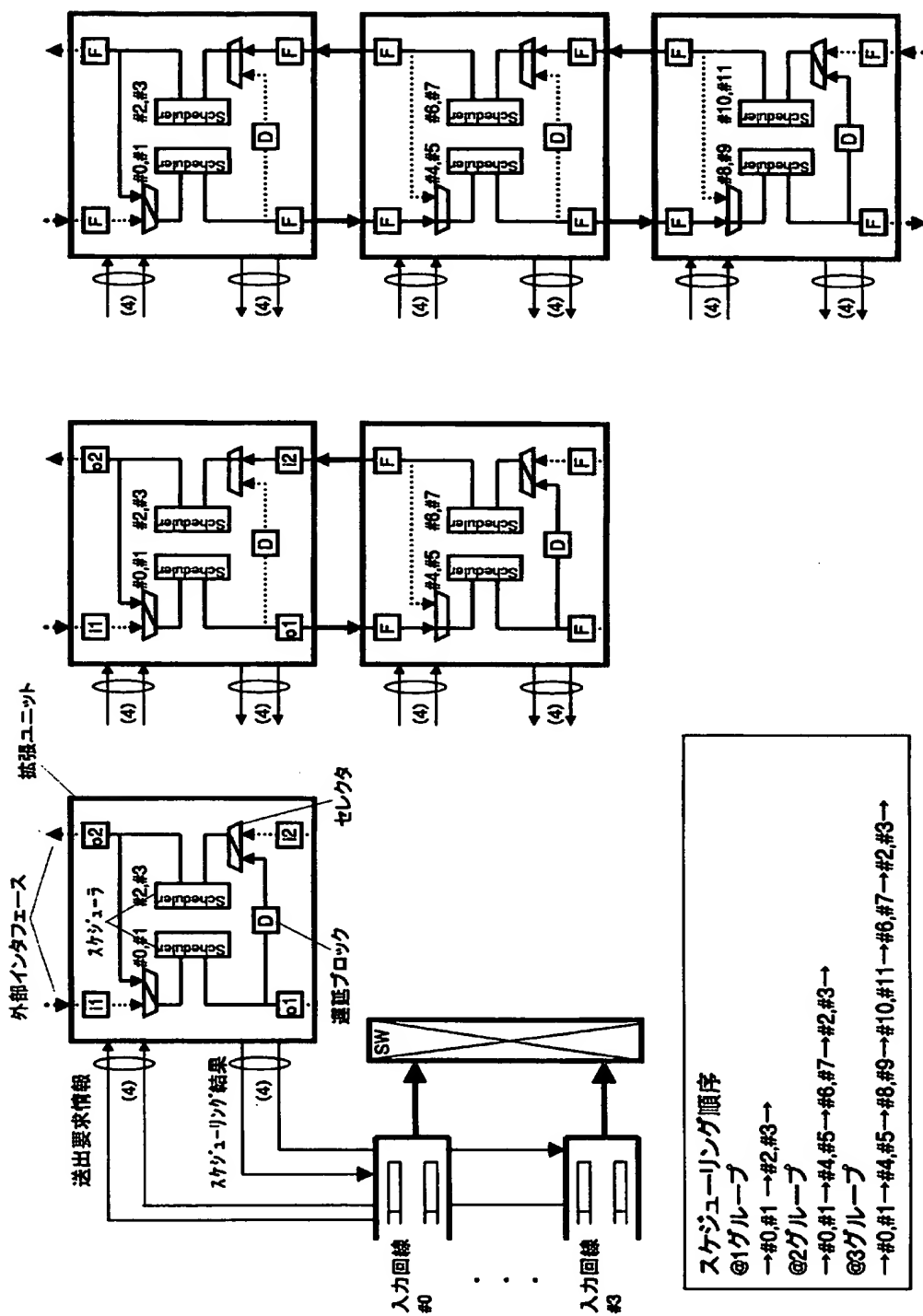
実施例 6 における分散スケジューラ型入カバツファスイッチを説明するための図

【図 3 2】



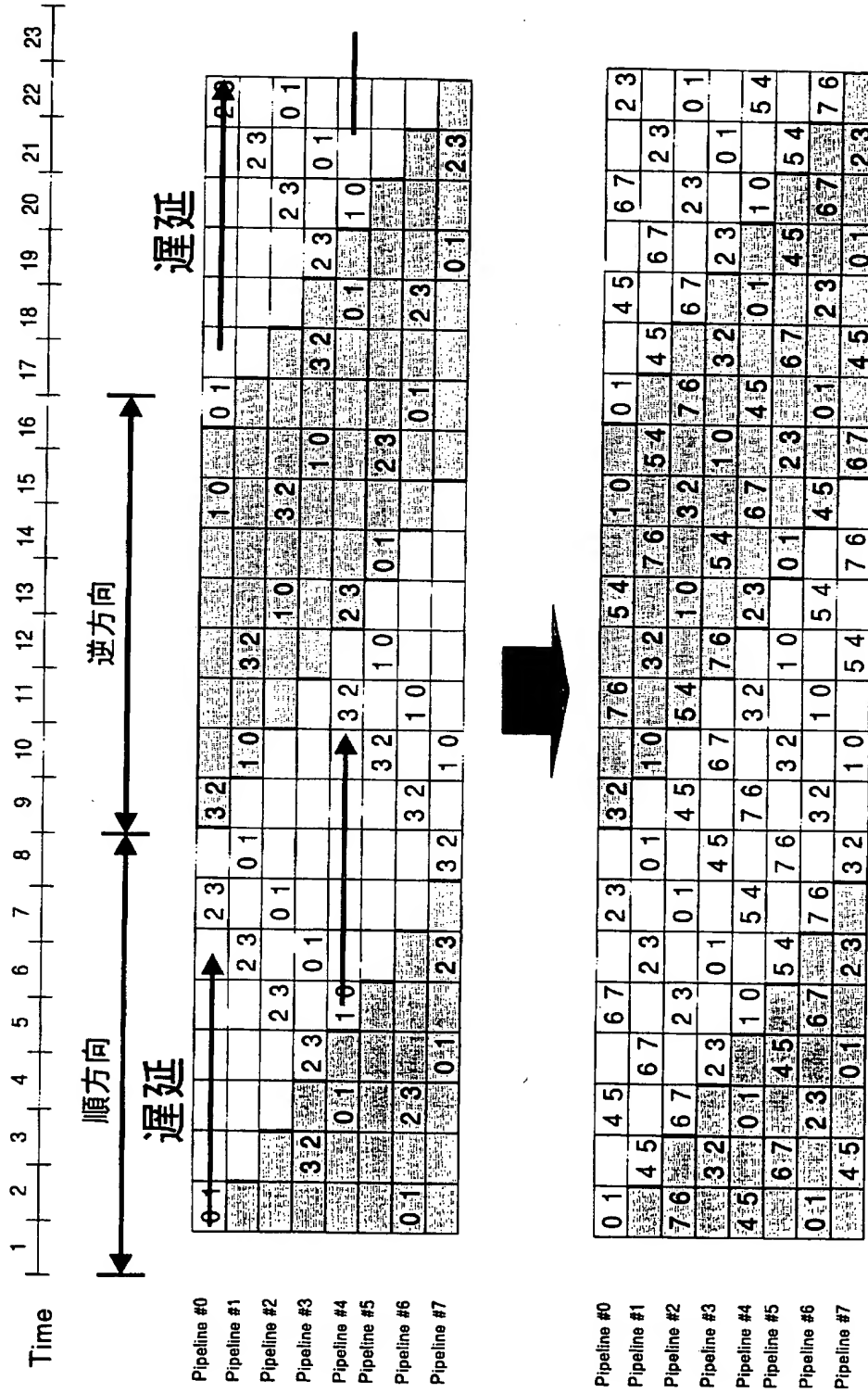
実施例 6 におけるスケジューラの拡張方法を説明するための図

【图 3 3】



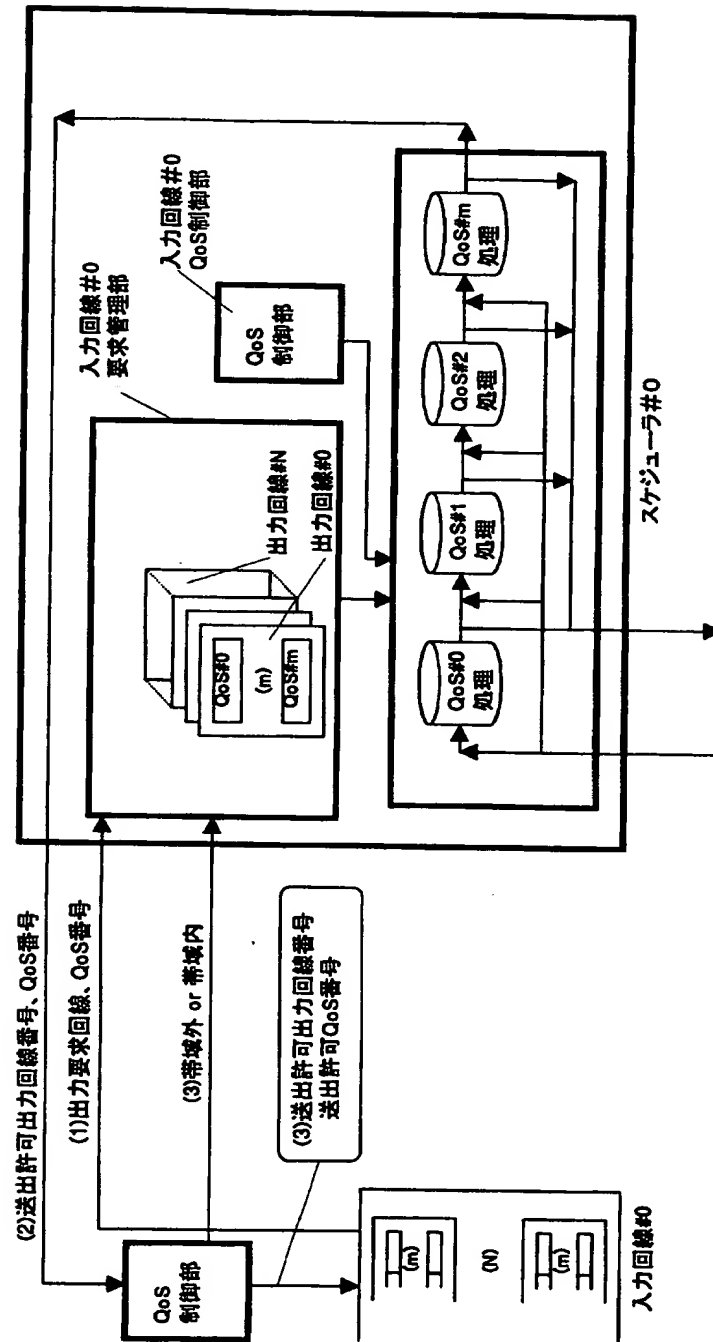
実施例6における分散スケジューラ拡張構成を説明するための図

【図 3 4】



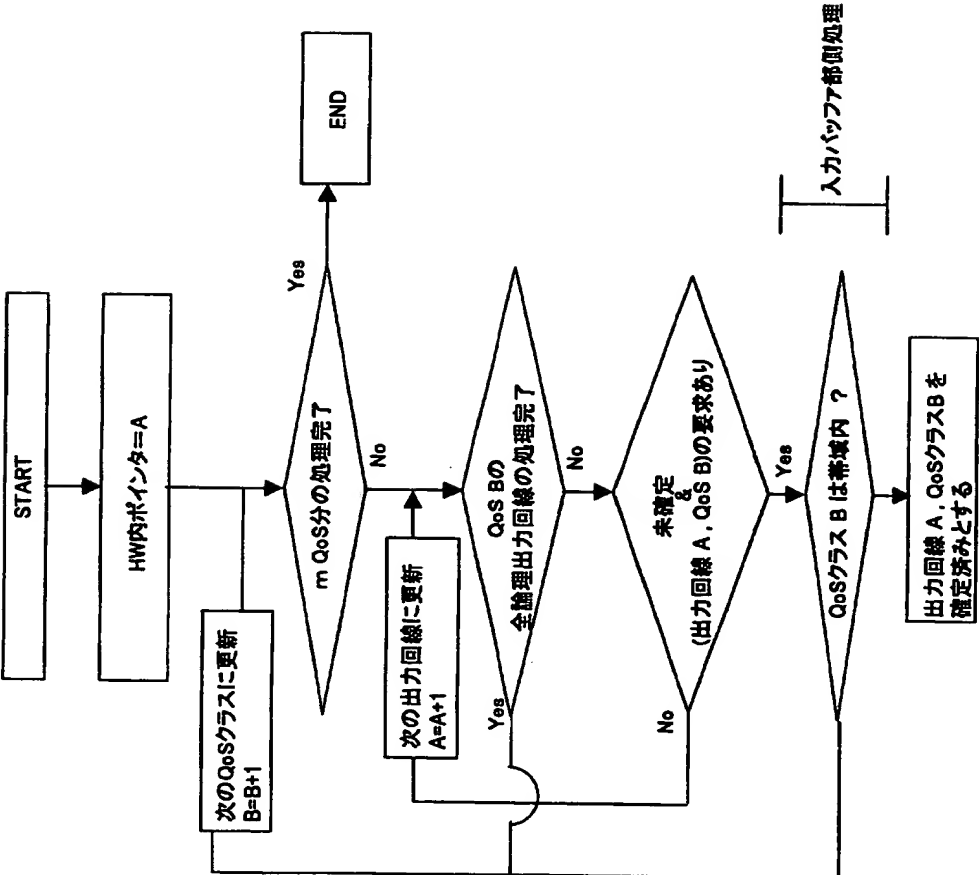
実施例 6 における遅延ブロックの動きを示す説明図

【図 35】



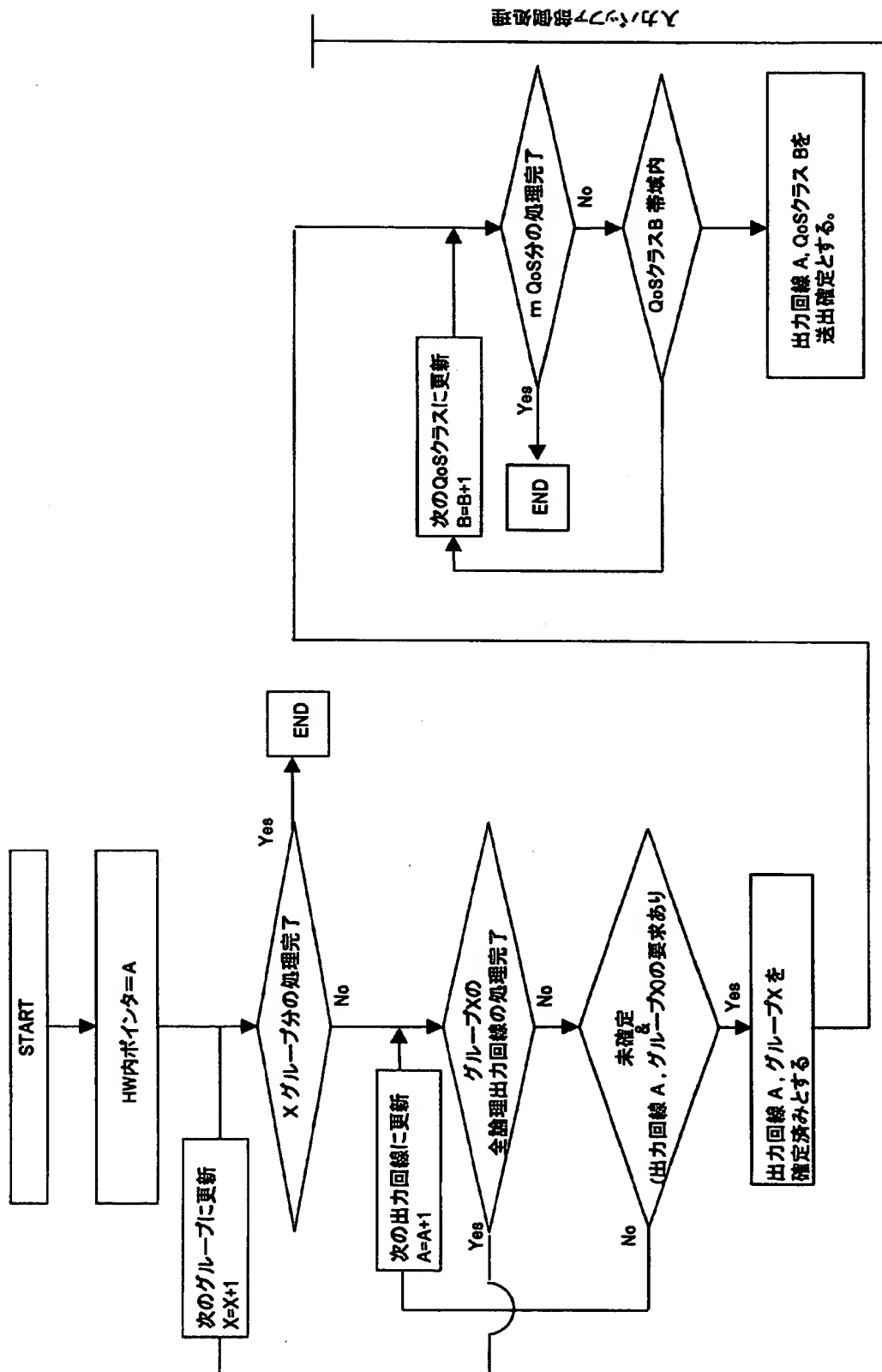
実施例 7 における QoS 制御のためのブロック図

【図 36】



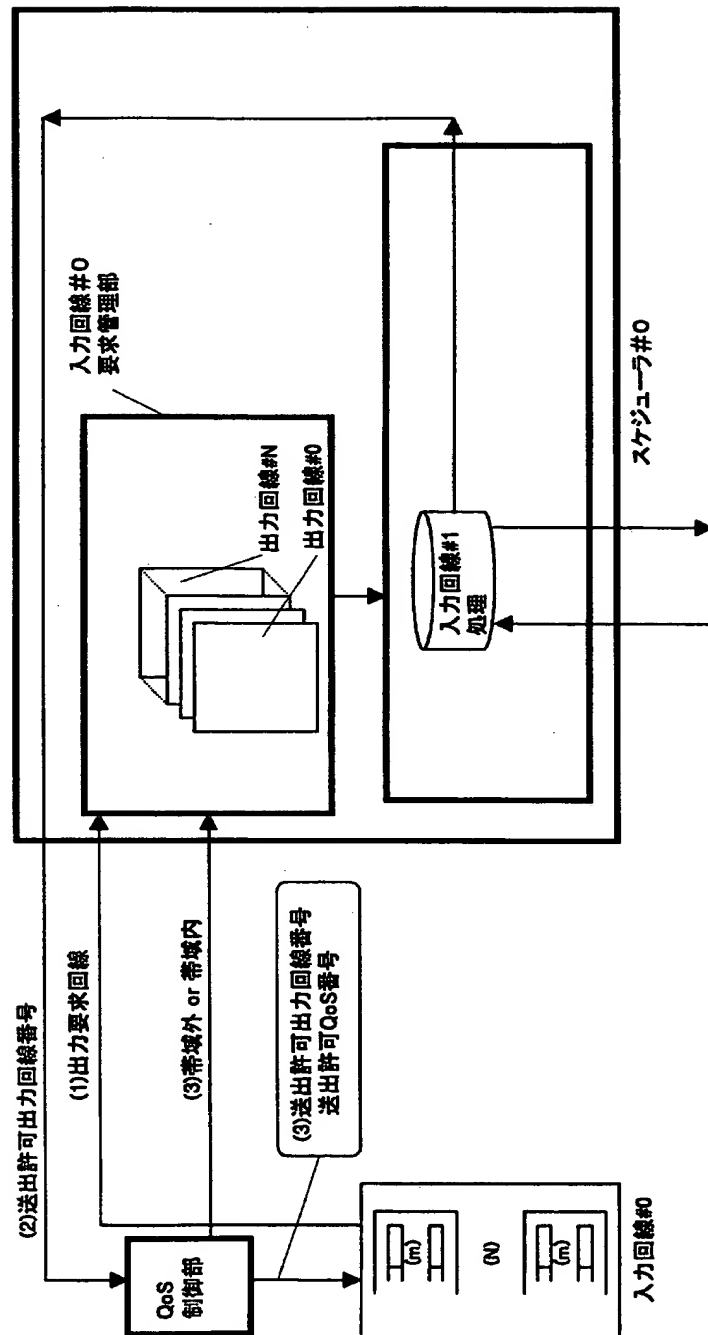
実施例 7 における QoS 制御のフロー図

【図 37】



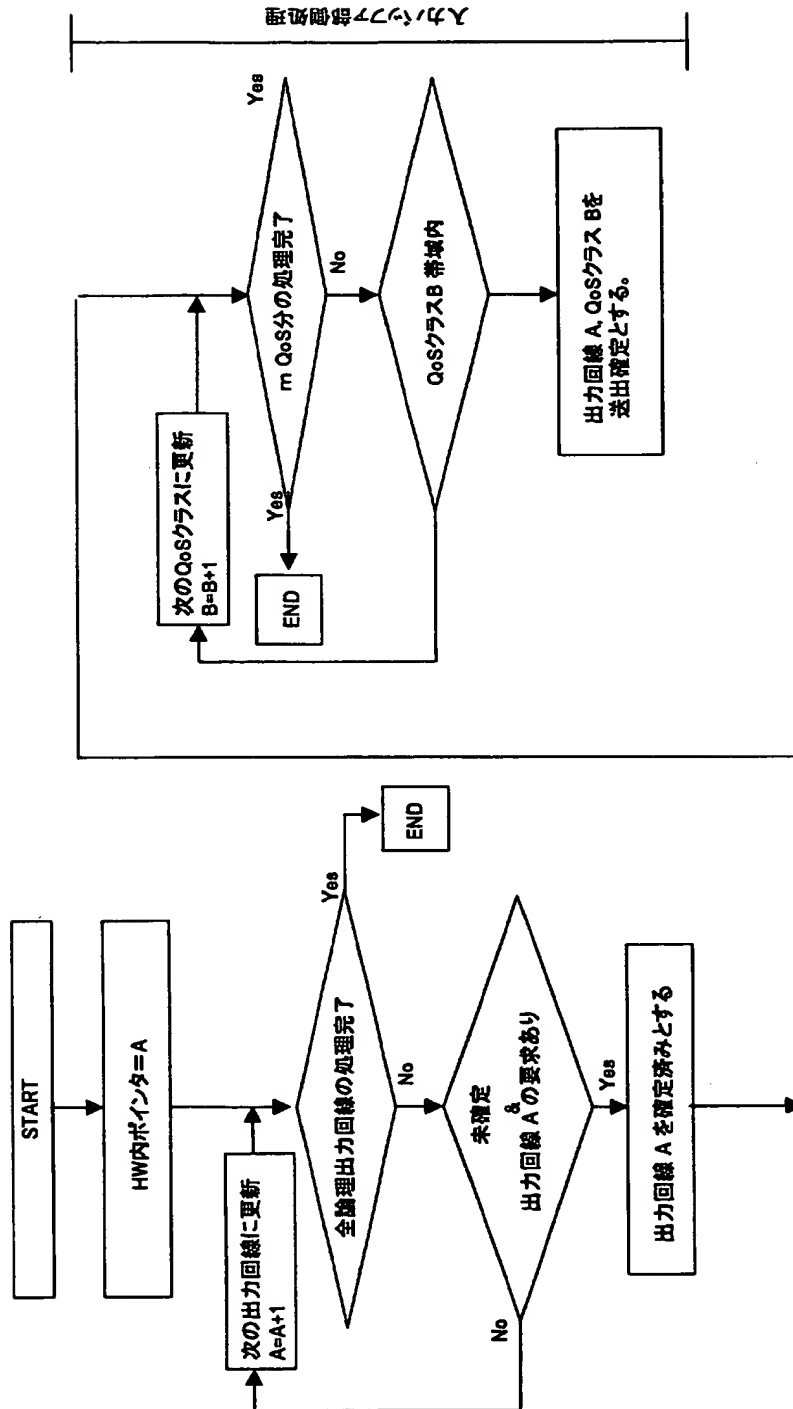
実施例 8 における QoS 制御のフロー図

【図 38】



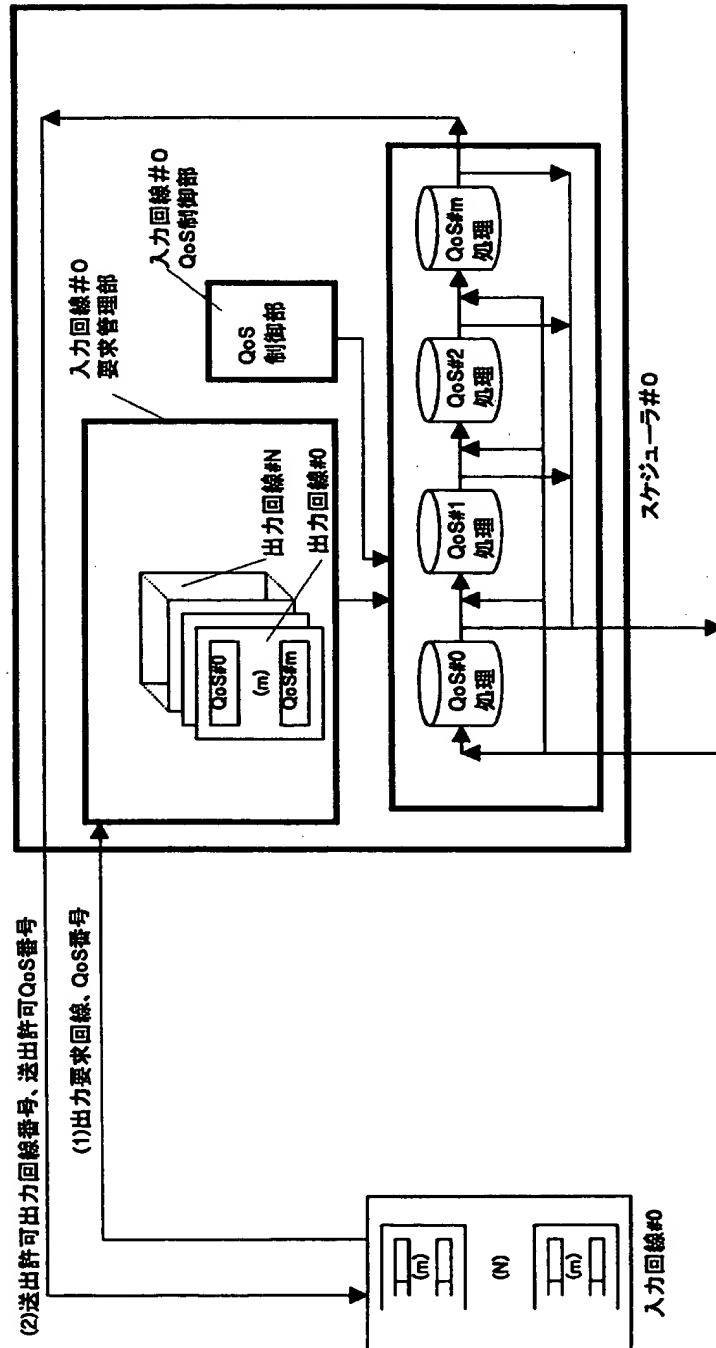
実施例 9 における QoS 制御のためのブロック図

【図 39】



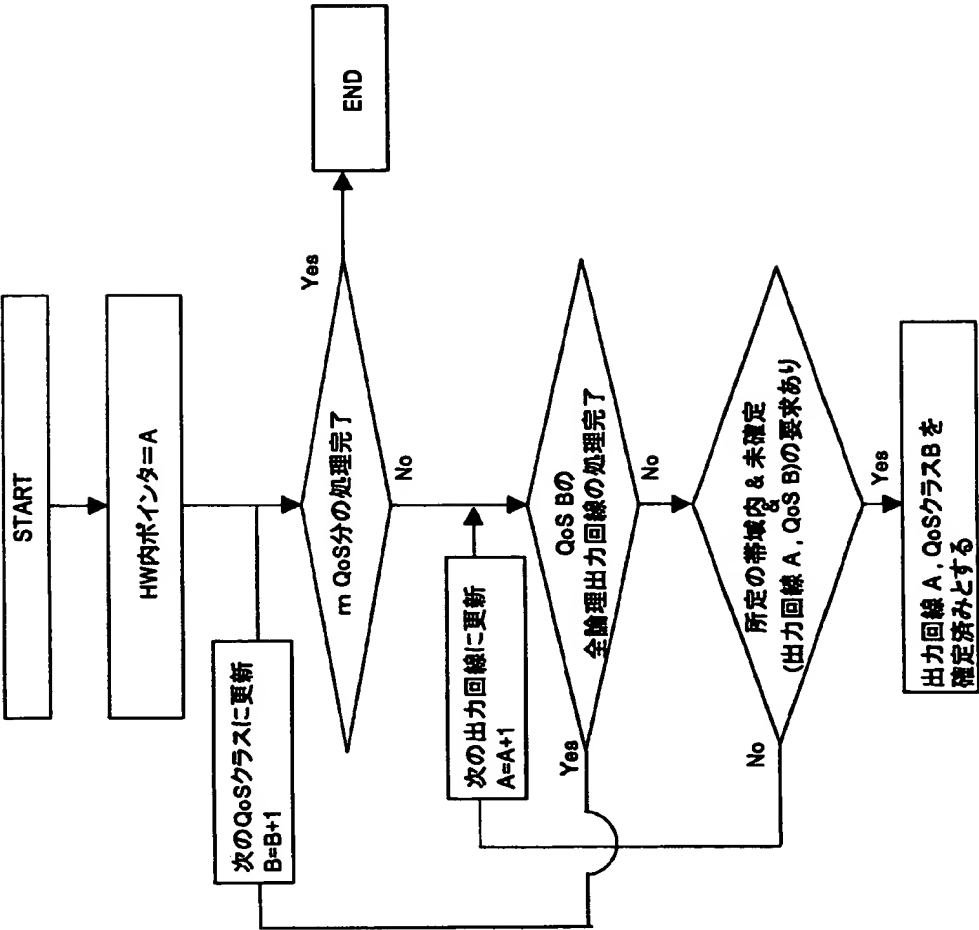
実施例 9 における QoS 制御のフロー図

【図 40】



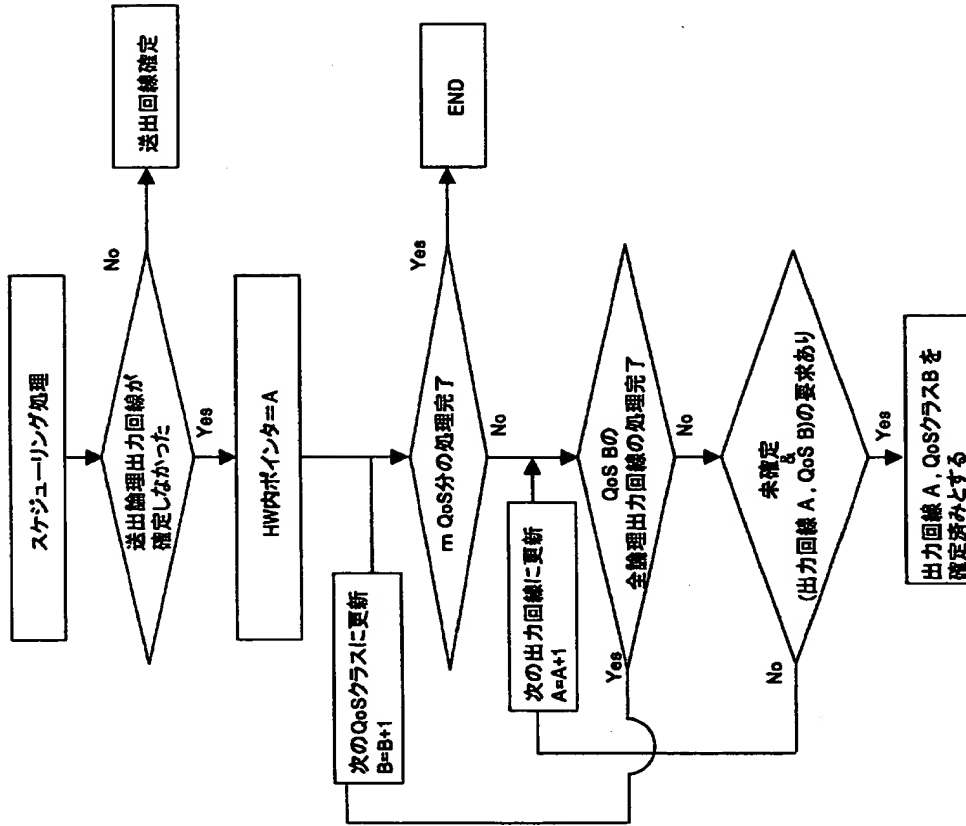
実施例 10 における QoS 制御のためのブロック図

【図 4 1】



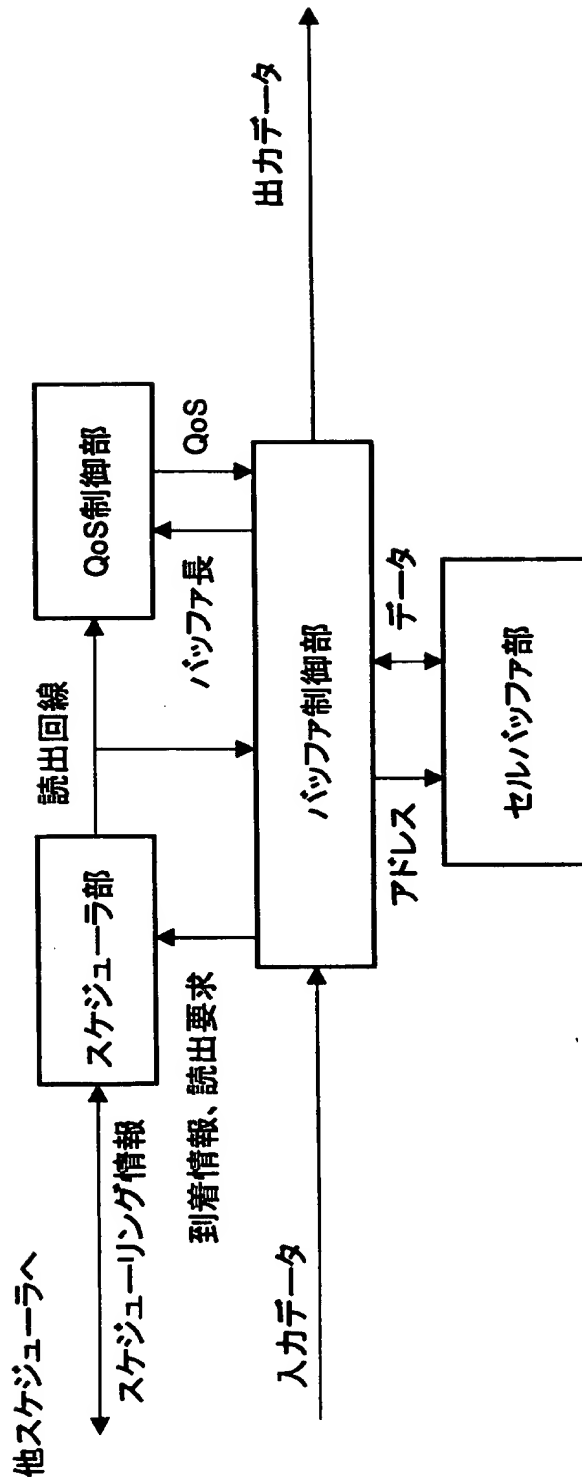
実施例 10 における帯域制御フロー図

【図 4 2】



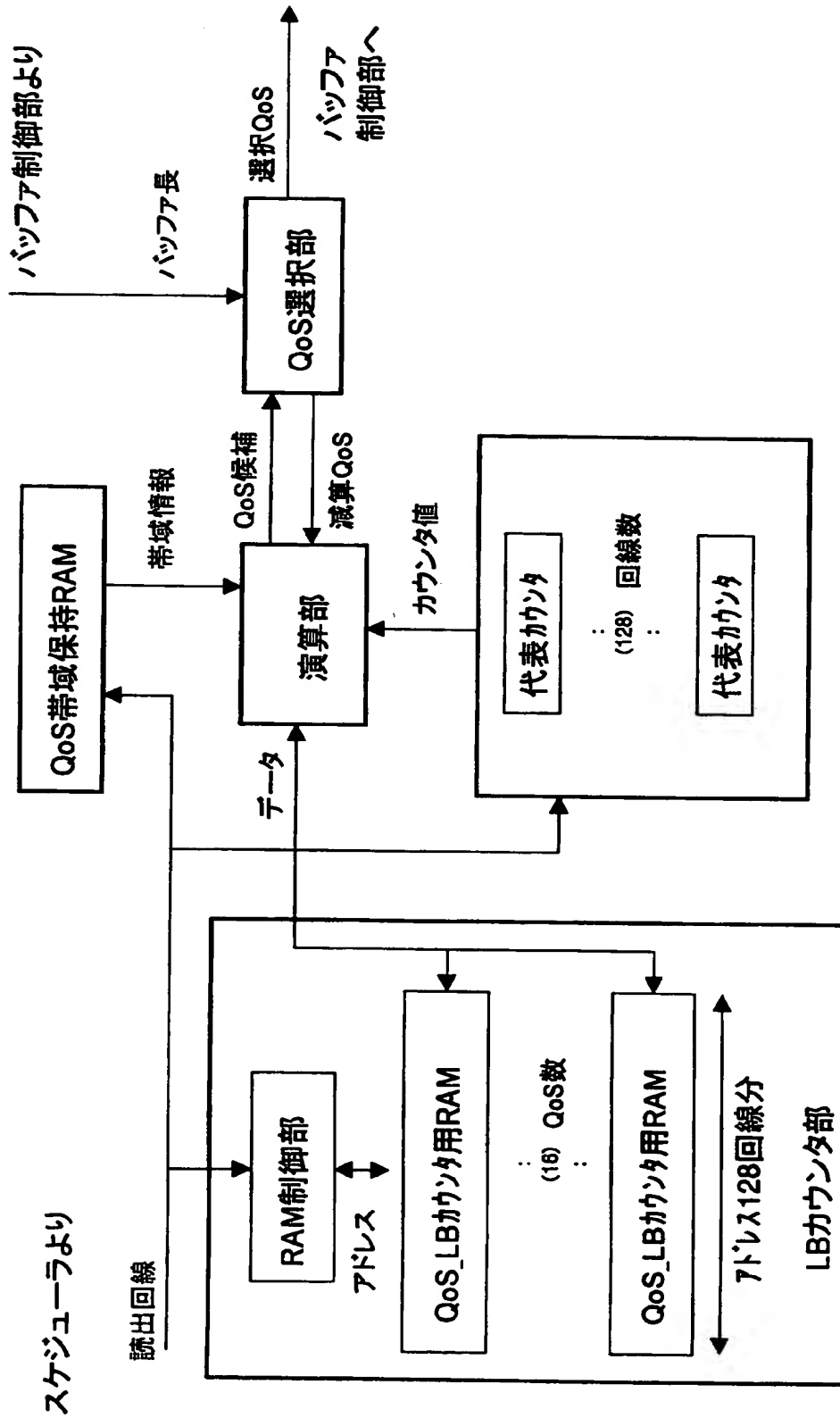
実施例 1 1 における帯域制御フロー図

【図 43】



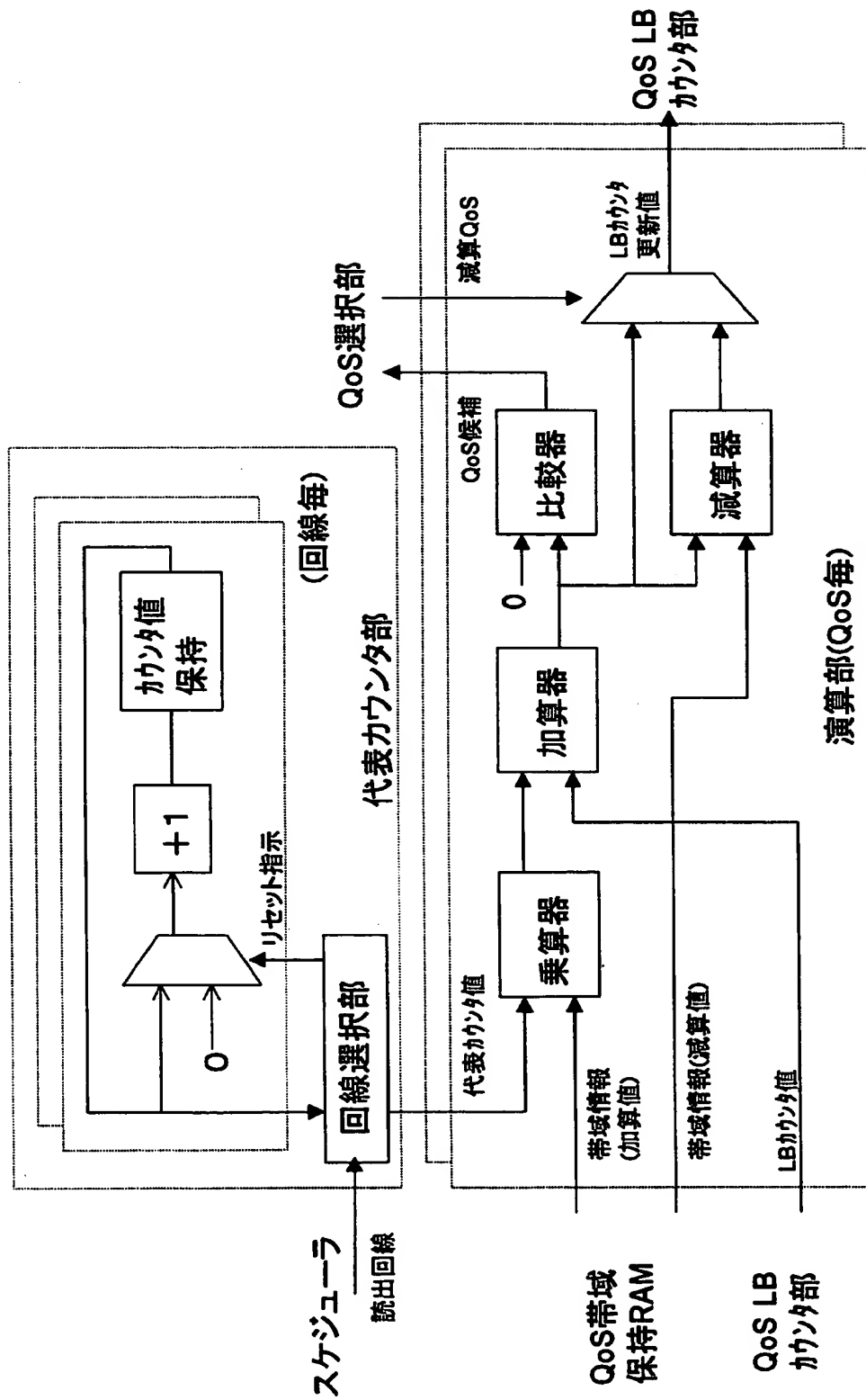
実施例 12 における入力バッファの構成図

【図 44】



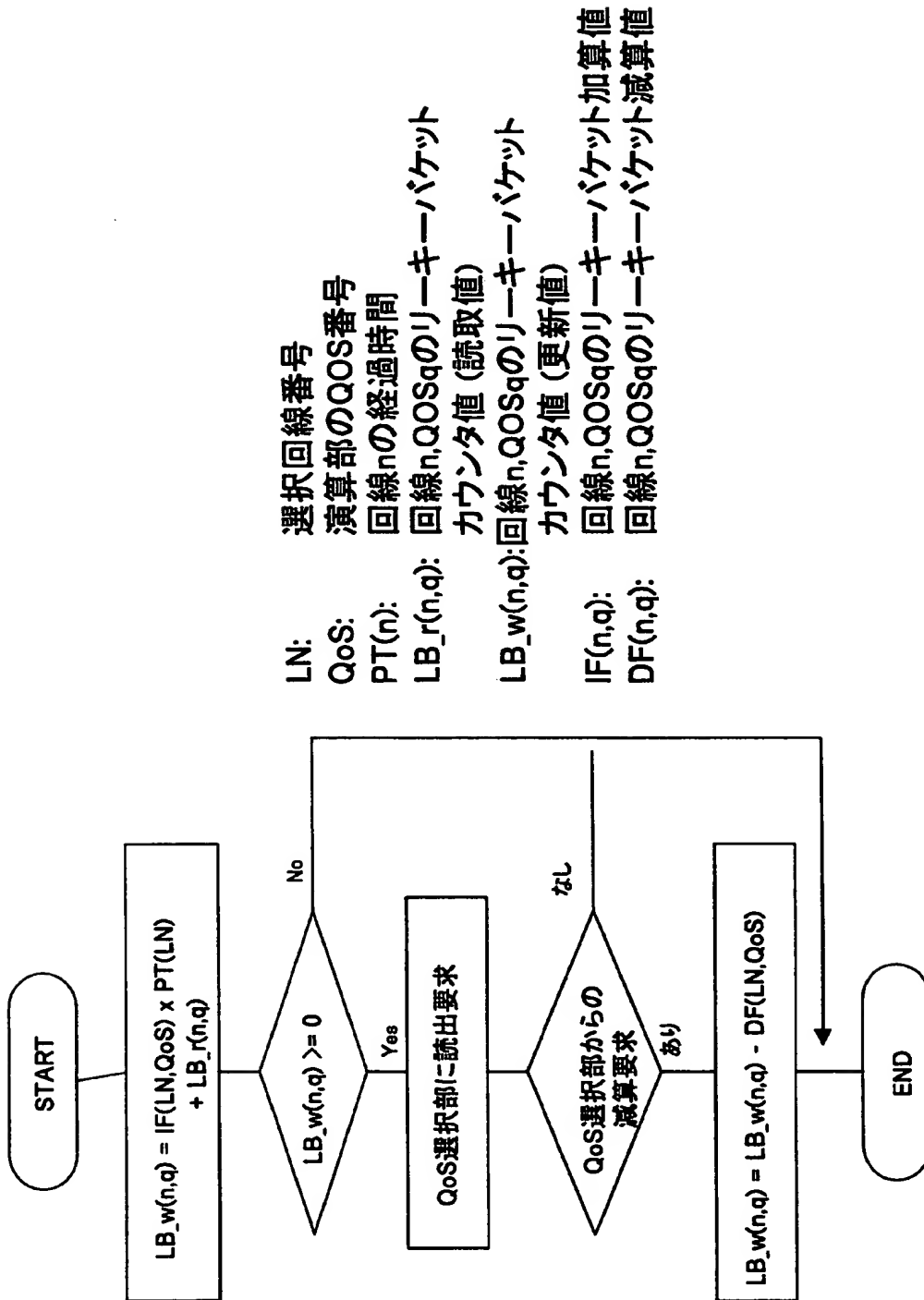
実施例 12における入カバッファのQoS制御部のブロック構成図

【図 45】



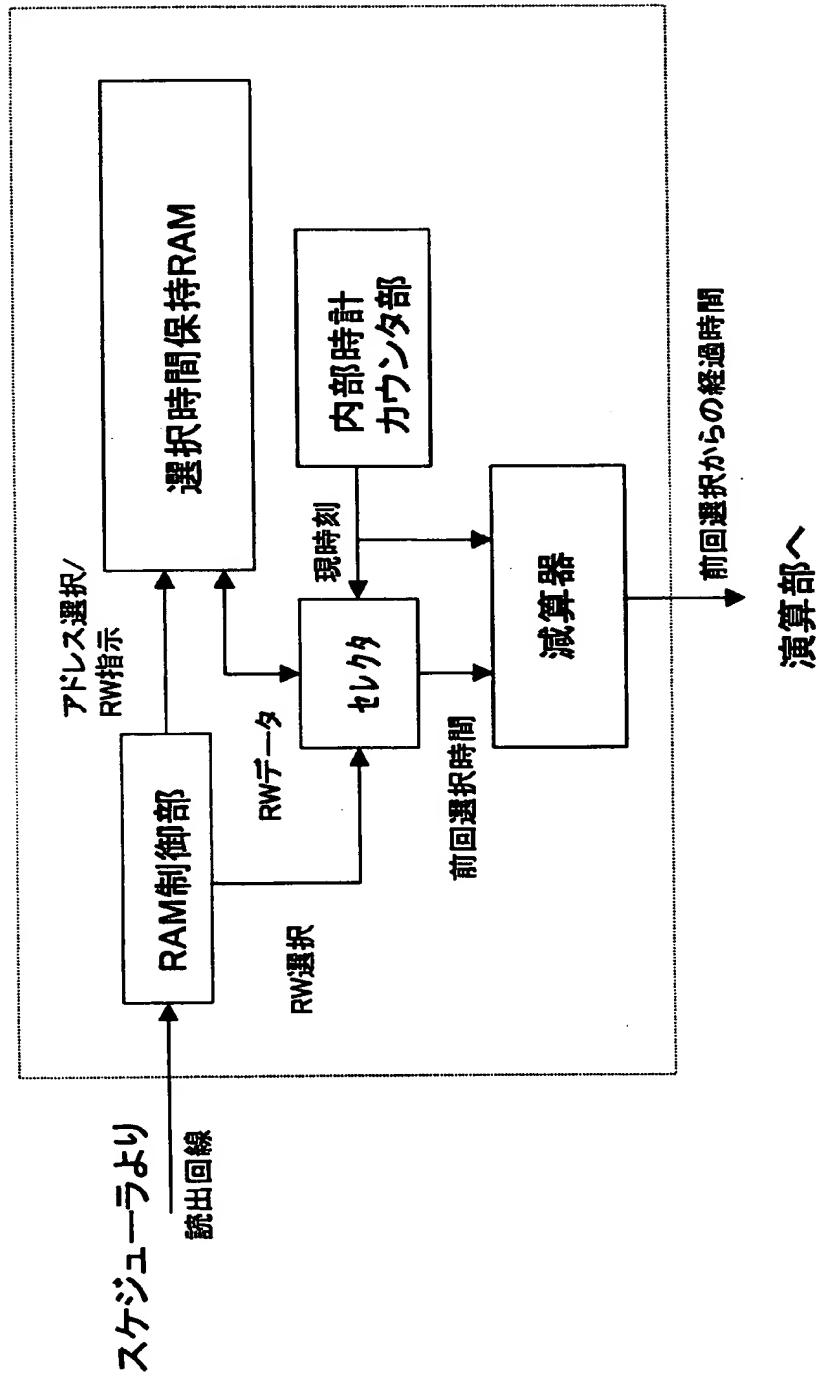
実施例 12における代表カウンタと演算部のブロック構成図

【図 4 6】



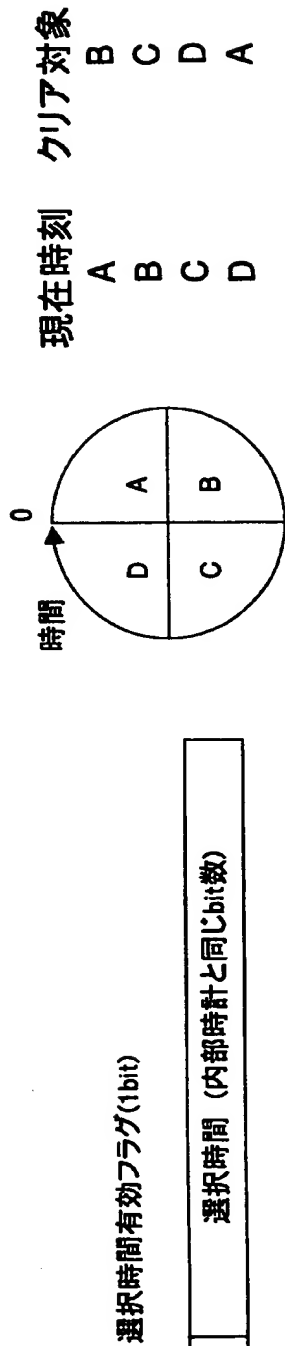
実施例 1 2 における演算部の動作アルゴリズムを示すフロー図

【図 47】

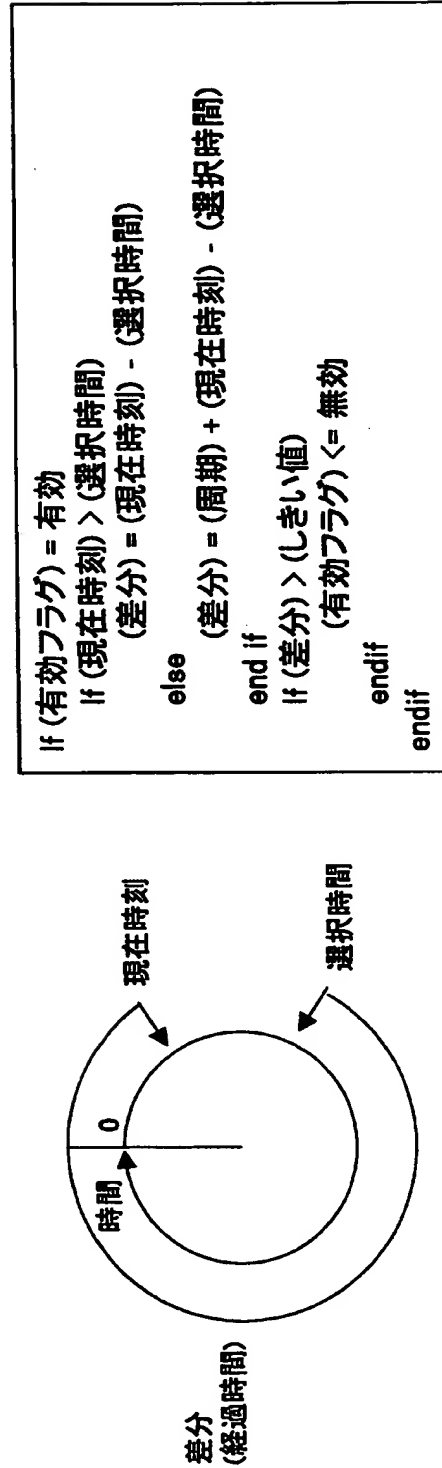


実施例 12 における代表カウンタの構成例を示す図

【図 4 8】



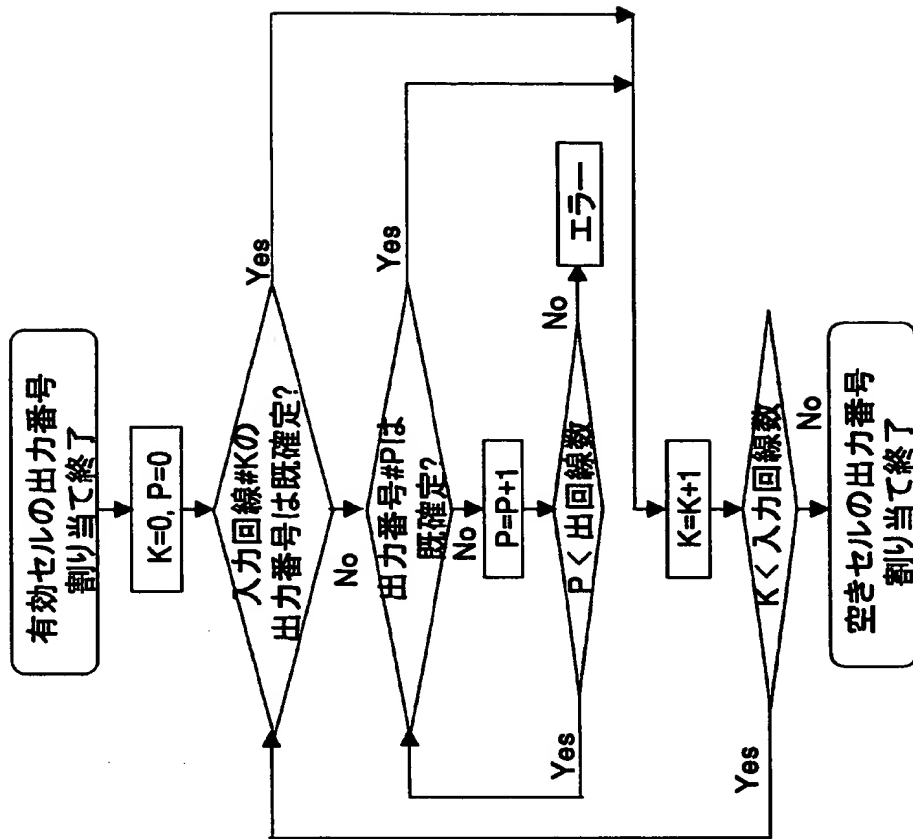
(a) 前回選択時間保持RAM ビットMAP(例) (b) 時間領域による有効時間チェック方式



(c) しきい値による有効時間チェック方式

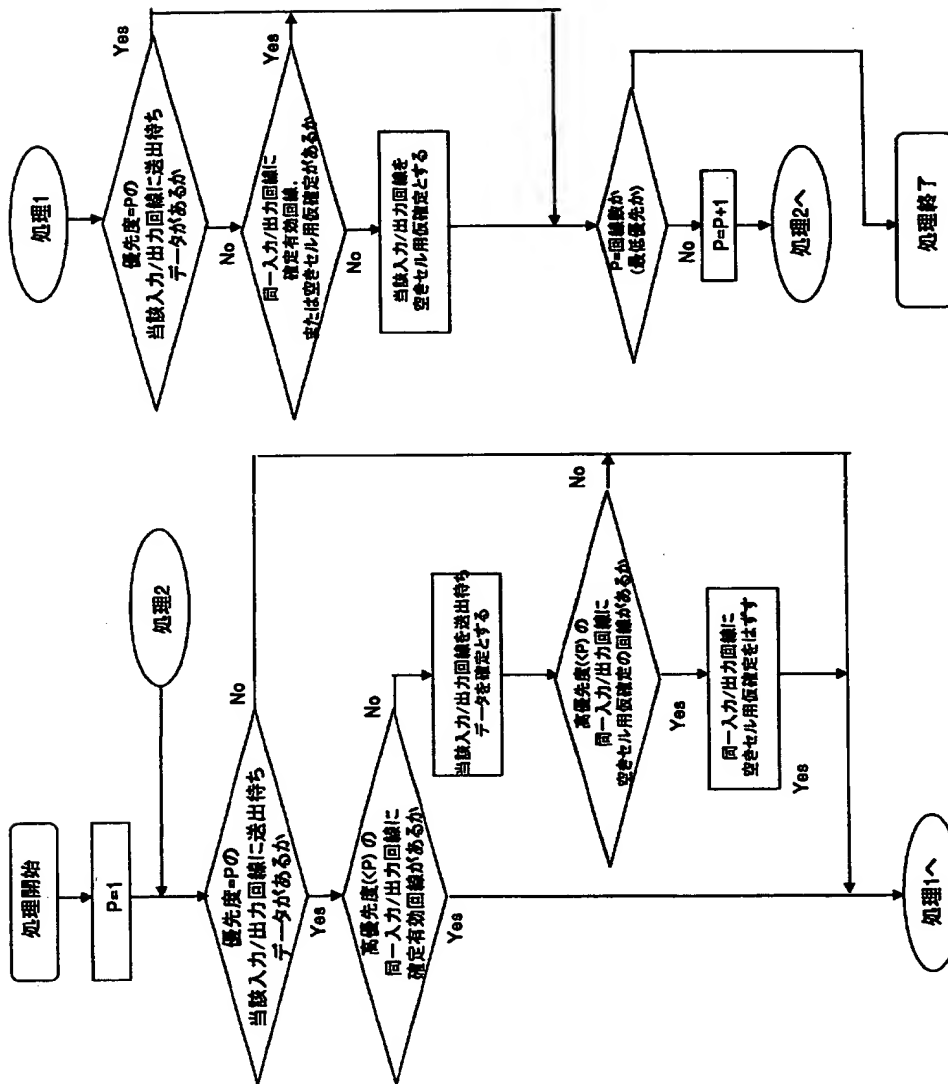
実施例 1 2 における選択時間の周期監視の説明

【図 49】



実施例 15 における処理フロー図

【図 50】



実施例 16 における処理フロー図

【書類名】 要約書

【要約】

【課題】 均等、不均等負荷のもとでも特性の劣化がなく、かつ、高速な繰り返しスケジューリングや複雑な演算処理を必要とせず、簡素でかつ処理速度がデバイス能力に依存しないスケジューリングシステムを実現する。

【解決手段】 ハイウェイ間ポインタを、全回線分のスケジューリングが終わった時点で、隣接する回線（右隣）に更新する。このとき、同一方向（時計回り）へN回更新された場合は、次のN回のスケジューリング処理は逆方向（反時計回り）の隣接回線へ更新する。

【選択図】 図1

【書類名】 職権訂正データ
【訂正書類】 特許願

<認定情報・付加情報>

【特許出願人】
【識別番号】 000005223
【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号
【氏名又は名称】 富士通株式会社
【代理人】 申請人
【識別番号】 100089244
【住所又は居所】 東京都中央区東日本橋3丁目4番10号 ヨコヤマ
ビル6階 秀和特許法律事務所
【氏名又は名称】 遠山 勉
【選任した代理人】
【識別番号】 100090516
【住所又は居所】 東京都中央区東日本橋3丁目4番10号 ヨコヤマ
ビル6階 秀和特許法律事務所
【氏名又は名称】 松倉 秀実

出 願 人 履 歴 情 報

識別番号 [000005223]

1. 変更年月日 1996年 3月26日

[変更理由] 住所変更

住 所 神奈川県川崎市中原区上小田中4丁目1番1号

氏 名 富士通株式会社